

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Query Match 68.5%; Score 150; DB 20; Length 221;
 Best Local Similarity 100.0%; Pred. No. 7.4e-135;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLRYRPNVTP 60
 DB 1 MGLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLRYRPNVTP 60
 QY 61 SYPGNTYDTGLPSYPMWLTSPGPPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 DB 61 SYPGNTYDTGLPSYPMWLTSPGPPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 QY 121 AAAPAPPIAAEPAAAPLTPTVAAEPAA 150
 DB 121 AAAPAPPIAAEPAAAPLTPTVAAEPAA 150

RESULT 7

AA30653
 ID AAM30653 standard; Protein; 219 AA.

AA30653;
 AC AAM30653;

DY 12-APR-1999 (first entry)

DB Human secreted protein clone cp16 1 protein.

KW Human; secreted protein; nutritional activity; cytokine; vaccine;
 KW cell proliferation; differentiation; immune stimulation; suppression;
 KW haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
 KW chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
 KW tumour invasion suppression; tumour inhibition.

OS Homo sapiens.

PN MO9901466-A1.

PD 14-JAN-1999.

PF 01-JUL-1998; 98MO-US13813.

PR 27-OCT-1997; 97US-0958304.

PR 02-JUL-1997; 97US-0887195.

(GEMT) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
 PI Spaulding V, Treacy M;

DR WPI; 1999-105994/09.

DR N-PSDB; AAB80740.

PT New polynucleotides encoding secreted human proteins - are derived
 PT from human foetal brain, adult testes, adult brain, foetal kidney,
 PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
 PT potential vaccines

PT Claim 24; Page 71-72; 107pp; English.

XX The present sequence represents a human secreted protein from clone
 CC cp16.1, deposited as ATCC 98482. Human secreted protein clone
 CC polynucleotides and proteins are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals. Suggested
 CC activities include nutritional activity, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating (e.g. as
 CC vaccines) or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
 CC suppressor activity, and tumour inhibition activity. The polynucleotides
 CC are also stated to be useful for gene therapy.

XX SQ Sequence 219 AA;
 Query Match 63.9%; Score 140; DB 20; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-125;
 Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLRYRPNVTP 60
 DB 1 MGLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLRYRPNVTP 60
 QY 61 SYPGNTYDTGLPSYPMWLTSPGPPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 DB 61 SYPGNTYDTGLPSYPMWLTSPGPPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 QY 121 AAAPAPPIAAEPAAAPLTPTVAAEPAA 140
 DB 121 AAAPAPPIAAEPAAAPLTPTVAAEPAA 140

RESULT 8

ABB64202
 ID ABB64202 standard; Protein; 75 AA.

AC ABB64202;

DT 26-MAR-2002 (first entry)

DB Drosophila melanogaster polypeptide SEQ ID NO 19398.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2001; 2000US-191637P.

PR 11-JUL-2000; 2000US-061415P.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EM;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL08305.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PT Disclosure; SEQ ID/NO 19398; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genetic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://wipo.int/pub/published_pct_sequences.

XX SQ Sequence 75 AA;
 Query Match 4.1%; Score 9; DB 22; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.91;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:22:36 ; Search time 41 Seconds
(without alignments)
847.832 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 219
Sequence: 1 MKLLIMACTVCVAFARRRR.....TAKDAPDPHPSPSLQANQ 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	100.0	219	22 AAB60109	Human transport pr
2	219	100.0	219	23 AAB27862	Human zslg63 prote
3	219	100.0	219	23 ABG31608	Human secreted sal
4	219	100.0	219	23 AAU74536	Human zslg63 polyp
5	219	100.0	219	24 ABU08515	Human zslg63 prote
6	150	68.5	221	20 AA19472	Amino acid sequenc
7	140	63.9	219	20 AA130653	Human secreted pro
8	9	4.1	75	22 ABB64202	Drosophila melanog
9	9	4.1	322	17 AAW03688	Leishmania chagasi

10	9	4.1	728	22 ABB65430	Drosophila melanog
11	9	4.1	1874	22 AAB76532	Corynebacterium g1
12	9	4.1	2969	22 AAG06680	C glutamicum prote
13	8	3.7	9	24 ABP66450	H1A class I molecu
14	8	3.7	59	24 ABP80080	N. gonorrhoeae ami
15	8	3.7	65	24 ABG27598	Arabidopsis thalia
16	8	3.7	82	21 AAG19327	Arabidopsis thalia
17	8	3.7	94	21 AAG19326	Arabidopsis thalia
18	8	3.7	106	21 AAG19325	Arabidopsis thalia
19	8	3.7	121	21 AAB40893	Human ORFX ORF57
20	8	3.7	124	21 AAG27837	Arabidopsis thalia
21	8	3.7	129	22 AAG6380	Amino acid sequenc
22	8	3.7	132	20 AAY34678	Chlamydia pneumon
23	8	3.7	139	21 AAG54857	Arabidopsis thalia
24	8	3.7	145	21 AAB32749	Eucalyptus grandis
25	8	3.7	145	21 AAG07525	Arabidopsis thalia
26	8	3.7	145	21 AAG45122	Arabidopsis thalia
27	8	3.7	145	21 AAG45119	Arabidopsis thalia
28	8	3.7	181	21 AAG06866	Arabidopsis thalia
29	8	3.7	181	21 AAG10820	Arabidopsis thalia
30	8	3.7	185	21 AAG54822	Arabidopsis thalia
31	8	3.7	193	21 AAG06865	Arabidopsis thalia
32	8	3.7	193	21 AAG10819	Arabidopsis thalia
33	8	3.7	205	21 AAG06864	Arabidopsis thalia
34	8	3.7	205	21 AAG10818	Arabidopsis thalia
35	8	3.7	249	24 ABR40639	Triticum aestivum
36	8	3.7	328	21 AAY90267	Protein chimera Q
37	8	3.7	372	22 AAG16669	Novel human diagno
38	8	3.7	412	21 AAY90266	Protein chimera O.
39	8	3.7	412	24 AAG73864	L. infantum anti ge
40	8	3.7	421	22 ABB67110	Drosophila melanog
41	8	3.7	428	23 AAB50963	Maize methyl CPg b
42	8	3.7	428	24 AAB70532	Histone deacetylase
43	8	3.7	493	22 AAB70532	Drosophila melanog
44	8	3.7	511	22 ABB68667	Human liver peptid
45	8	3.7	511	22 ABB28667	Peptide #1318 enco

ALIGNMENTS

RESULT 1	
AA60109	standard; Protein; 219 AA.
XX	AA60109;
AC	28-MAR-2001 (first entry)
DT	Human transport protein TPPT-29.
DE	Human; transport protein; TPPT; transport disorder; metabolic disorder;
XX	neurological disorder; cardiovascular disorder; reproductive disorder;
KW	Immune disorder; cancer.
KW	Homo sapiens.
XX	WO200078953-A2.
PN	28-DEC-2000.
PD	16-JUN-2000; 2000WO-US16668.
PF	17-JUN-1999; 99US-0139923.
XX	10-AUG-1999; 99US-0148177.
PR	18-AUG-1999; 99US-0149357.
XX	28-OCT-1999; 99US-0162287.
PA	(INCYTE-) INCYTE GENOMICS INC.
XX	LaI P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI	Baughn MR, Azimzai Y, Lu DM, Au-Young J, Patterson C;
XX	

DR WPI; 2001-041424/05.
 DR N-PSDB; AAE27729.
 XX
 PT Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 XX
 PS Claim 2; Page 130; 165pp; English.
 XX
 CC The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated TPTs). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 CC
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 22; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1,1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLIACIVCAVAPARRRRPFGEDNDGHLHPSLNPYGRRLPPLYYRPVNTVP 60
 DB 1 MKLLIACIVCAVAPARRRRPFGEDNDGHLHPSLNPYGRRLPPLYYRPVNTVP 60
 QY 61 SYRGNTYTDGLSPYMLTSPGPPYVYHIGFPLATQLVNPPRGPFPVPSRPFSA 120
 DB 61 SYRGNTYTDGLSPYMLTSPGPPYVYHIGFPLATQLVNPPRGPFPVPSRPFSA 120
 QY 121 AAAPAAPPIAEPAAAAPLTATPVAAEPAAAGAVAAEPAAEPAAEPAAEPAA 180
 DB 121 AAAPAAPPIAEPAAAAPLTATPVAAEPAAAGAVAAEPAAEPAAEPAAEPAA 180
 QY 181 EAPVGVPAEAPSPAPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219
 DB 181 EAPVGVPAEAPSPAPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219
 RESULT 2
 AAE27862
 ID AAE27862 standard; Protein; 219 AA.
 XX
 AC AAE27862;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 DE Human zsig63 protein.
 XX
 XX Human; secreted salivary protein; zsig63 protein; host defense protein;
 KM immune modulating factor; antipathogenic; cell-cell signaling molecule;
 KM growth factor; cytokine; growth factor hormone activity; dental caries;
 KM infection; tooth decay; periodontal disease; gastrointestinal disease;
 KM through; urinary tract infection; vaginal infection; diabetes; obesity;
 KM anti-inflammatory; chronic tissue damage; lung dysfunction; restenosis;
 KM gene therapy; salivary gland dysfunction; prostate gland dysfunction;
 KM forensic DNA profiling; chondrosarcoma; atherosclerosis; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT 1..15
 FT /label= "Signal_peptide"
 FT 16..219
 FT /note= "Mature human zsig63 protein"
 FT 14..21
 FT /note= "Hydrophilic region"
 FT 16..37
 FT /note= "Domain 1"
 FT 17..33
 FT /note= "Antigenic epitope"
 FT 24..30
 FT /note= "Hydrophilic region"
 FT 38..126
 FT Domain

FT /note= "Domain 2"
 FT 66..73
 FT /note= "Antigenic epitope"
 FT 103..108
 FT /note= "Antigenic epitope"
 FT 124..133
 FT /note= "Repeat 1"
 FT 127..219
 FT /note= "Domain 3"
 FT 134..138
 FT /note= "Repeat 2"
 FT 139..143
 FT /note= "Repeat 3"
 FT 144..148
 FT /note= "Repeat 4"
 FT 149..153
 FT /note= "Repeat 5"
 FT 154..158
 FT /note= "Repeat 6"
 FT 159..163
 FT /note= "Repeat 7"
 FT 164..168
 FT /note= "Repeat 8"
 FT 169..173
 FT /note= "Repeat 9"
 FT 174..178
 FT /note= "Repeat 10"
 FT 179..183
 FT /note= "Repeat 11"
 FT 184..188
 FT /note= "Repeat 12"
 FT 187..192
 FT /note= "Hydrophilic region"
 FT 189..193
 FT /note= "Repeat 13"
 FT 190..197
 FT /note= "Antigenic epitope"
 FT 194..198
 FT /note= "Repeat 14"
 FT 199..203
 FT /note= "Repeat 15"
 FT 202..215
 FT /note= "Antigenic epitope"
 FT 204..208
 FT /note= "Repeat 16"
 XX
 XX US2002090677-A1.
 XX
 XX 11-JUL-2002.
 XX
 PD 03-AUG-2001; 2001US-0923236.
 XX
 PP 17-MAR-1999; 99US-124820P.
 PR 17-MAR-2000; 2000US-0527345.
 XX
 XX (ADLER/) ADLER D A.
 PA (SHEP/) SHEPPARD P O.
 XX
 XX
 XX Adler DA, Sheppard PO;
 PI
 XX WPI; 2002-642378/69.
 XX
 DR N-PSDB; AAD45050.
 XX
 PT Novel secreted salivary polypeptide, zsig63, useful as antimicrobial
 PT agent for treating microbial infection, dental caries, periodontal
 PT disease, through gastrointestinal disease, and for aiding digestion -
 XX
 PS Claim 10; Page 28; 33pp; English.
 XX
 CC The invention relates to human secreted salivary polypeptide designated
 CC as zsig63 and nucleic acid molecules encoding such polypeptides. zsig63
 CC can be used in detecting agonists and antagonists of its activity, and
 CC is also useful as a host defense polypeptide, immune modulating factor,

CC antipathogenic polypeptide, cell-cell signalling molecule, growth factor,
 CC cytokine, or as secreted extracellular matrix associated proteins with
 CC growth factor hormone activity. It is useful for treating conditions
 CC associated with pathological microbes, including bacterial, fungal and
 CC viral infections, for treating dental caries (tooth decay), periodontal
 CC disease, thrush and gastrointestinal diseases, for treating urinary tract
 CC infection, vaginal infection and for preventing infection in skin and
 CC other epithelial wounds. zsig63 is useful for establishing normal
 CC microflora and protect against pathogenic colonisation and invasion, for
 CC treating chronic tissue damage e.g. damage in extremities associated
 CC with diabetes and useful as anti-inflammatory agents. It is useful as a
 CC marker of lung dysfunction, salivary gland dysfunction, or dysfunction of
 CC prostate gland. It is also therapeutically useful for aiding digestion.
 CC Polymucleotides of the invention are used in gene therapy for increasing
 CC or inhibiting zsig63 activity, for detecting abnormalities on human
 CC chromosome 4 associated with disease or other human traits and as
 CC diagnostics in forensic DNA profiling. Sequences of the invention are
 CC useful for stimulating proliferation or differentiation of cardiac
 CC myocytes, for proliferation or differentiation of adipocytes and for
 CC inhibiting chondrosarcomas, atherosclerosis, restenosis and obesity.
 CC The present sequence is human zsig63 protein. zsig63 gene is located
 CC on chromosome 4.

SO Sequence 219 AA;

Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLIMACTVCVAPARRRRPFPGEDDNDGHPHPSINITYGIRNLPPPLYRRPVNTVP 60
 Db 1 MKLLIMACTVCVAPARRRRPFPGEDDNDGHPHPSINIPYGINLPPLYYRRVNTVP 60
 Qy 61 SYPGNTYDTGSPWILTSRPFYVHNGFPLATOLNVPPLRGGPFVPPSPRPSA 120
 Db 61 SYPGNTYDTGSPWILTSRPFYVHNGFPLATOLNVPPLRGGPFVPPSPRPSA 120
 Qy 121 AAAPAAPPIAEPAAAPLTAIPVAAEPDAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 Db 121 AAAPAAPPIAEPAAAPLTAIPVAAEPDAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 Qy 181 EAPVGEPAEPEPAPAPATAKPAAPBPPSPSLQANQ 219
 Db 181 EAPVGEPAEPEPAPAPATAKPAAPBPPSPSLQANQ 219

RESULT 3

ABG31608 ID ABG31608 standard; Protein; 219 AA.

XX ABG31608;

DT 15-NOV-2002 (first entry)

DE Human secreted salivary protein zsig63.

KW Human; secreted salivary protein; zsig63; immunogen; zsig63-cytokine;
 KW antibody-cytokine; in vivo killing; pathological microbe; bacteria;
 KW fungal; viral; infection; salivary gland; anti-microbial; dental caries;
 KW tooth decay; periodontal disease; thrush; gastrointestinal disease;
 KW urinary tract infection; vaginal infection; skin infection; microflora;
 KW epithelial wound; pathogenic colonisation; invasion; pro-inflammatory;
 KW chronic tissue damage; vascular system; diabetes; anti-inflammatory;
 KW incompetent immune system; AIDS; acquired immunodeficiency syndrome;
 KW chemotherapy; radiation treatment; lung infection; cystic fibrosis;
 KW digestion; chromosome 4.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..15 /note= "Fusion protein peptide, specifically claimed
 FT in claim 18"

FT Region 1..219
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 9..204
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 14..19
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..21
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..37
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 17..33
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 24..29
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 24..33
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 25..30
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 38..126
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 66..73
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 103..108
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 109..215
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 124..133
 FT /label= Repeat_1
 FT 127..219
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 134..138
 FT /label= Repeat_2
 FT 139..143
 FT /label= Repeat_3
 FT 144..148
 FT /label= Repeat_4
 FT 149..153
 FT /label= Repeat_5
 FT 154..158
 FT /label= Repeat_6
 FT 159..163
 FT /label= Repeat_7
 FT 164..168
 FT /label= Repeat_8
 FT 169..173
 FT /label= Repeat_9
 FT 174..178
 FT /label= Repeat_10
 FT 179..183
 FT /label= Repeat_11
 FT 184..188
 FT /label= Repeat_12
 FT 187..192
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 189..193

FT Region /label= Repeat_13
 FT 190..197 /note="Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 194..198 /label= Repeat_14
 FT 199..203 /label= Repeat_15
 FT Region 202..215 /note="Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 204..208 /label= Repeat_16
 FT US2002081701-A1.
 XX 27-JUN-2002.
 PD 03-AUG-2001; 2001US-0922480.
 XX 17-MAR-1999; 99US-124820P.
 PR 17-MAR-2000; 2000US-0527345.
 XX (ADLER/) ADLER D A.
 PA (SHEP/) SHEPPARD P O.
 XX Adler DA, Sheppard PO;
 PI MPI: 2002-635468/68.
 DR N-PSDB; AHS2633, AHS2634.
 XX Novel secreted salivary protein, zsig63 and polynucleotide encoding it
 PT useful for treating microbial infections, inflammatory conditions,
 PT dental caries and lung infections associated with cystic fibrosis -
 XX
 XX Claim 10; Page 28; 33pp; English.
 CC The present invention relates to a new secreted salivary protein, zsig63.
 CC an antagonist or agonist of zsig63 protein activity. The invention is
 CC also useful as an immunogen for producing an antibody to zsig63
 CC polypeptide. zsig63-cytokine fusion proteins or antibody-cytokine fusion
 CC protein are useful for enhancing in vivo killing of target tissues.
 CC Pharmaceutical composition comprising purified zsig63 polypeptide are
 CC useful in the treatment of conditions associated with pathological
 CC microbes, including bacterial, fungal and viral infections. High
 CC expression of zsig63 in salivary gland suggests that anti-microbial
 CC polypeptides are useful for treatment of dental caries (tooth decay),
 CC periodontal disease, thrush and gastrointestinal disease. Other
 CC applications can be used in urinary tract infections, vaginal infections,
 CC prevention of infection in skin and other epithelial wounds. The
 CC polypeptides can be used to establish normal microflora and protect
 CC against pathogenic colonization and invasion. The invention is useful
 CC when pro-inflammatory activity is desired. Applications for
 CC such pro-inflammatory activity include the treatment of chronic tissue
 CC damage, particularly in areas having a limited or damaged vascular system
 CC e.g., damage in extremities associated with diabetes. Antagonists to
 CC zsig63 polypeptides may be useful as anti-inflammatory agents. The
 CC invention is useful for the treatment of patients having incompetent
 CC immune system, such as AIDS (acquired immunodeficiency syndrome) patients
 CC or individuals that have undergone chemotherapy, radiation treatment. The
 CC invention is also useful for the treatment of lung infections associated
 CC with cystic fibrosis and its agonists or antagonists are useful for
 CC aiding digestion. The present amino acid sequence represents the human
 CC secreted salivary protein zsig63 of the invention. This sequence is
 CC encoded by the human zsig63 gene located on chromosome 4.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACTVCVAFARKRRFPFPGDDDDGGHPLPSINTPYGINLLPPLLYRRPVNTVP 60
 DB 1 MKLLIMACTVCVAFARKRRFPFPGDDDDGGHPLPSINTPYGINLLPPLLYRRPVNTVP 60
 QY 61 STPGNTTYDTGCLPSFWILTSGFPYVYHIRGFPLATQLVNPPPLPRGPFVPSRPFSA 120
 DB 61 STPGNTTYDTGCLPSFWILTSGFPYVYHIRGFPLATQLVNPPPLPRGPFVPSRPFSA 120
 QY 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPVAAEPVAAEPVAAEPVAAEPVAAEPVAA 180
 DB 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPVAAEPVAAEPVAAEPVAAEPVAAEPVAA 180
 QY 181 BAPVGEPAABEPSPABEPATAPAAPEBPSPSLBOAQ 219
 DB 181 BAPVGEPAABEPSPABEPATAPAAPEBPSPSLBOAQ 219
 RESULT 4
 ID AAU74536 standard; Protein; 219 AA.
 XX AAU74536;
 AC 23-APR-2002 (first entry)
 DT Human zsig63 polypeptide.
 XX
 XX Human; zsig63; chromosome 4q12-q413; salivary protein; antimicrobial;
 KW microbial infection; tooth decay; periodontal disease; thrush; emphysema;
 KW gastrointestinal disease; urinary tract infection; vaginal infection;
 KW skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
 KW acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
 KW chronic bronchitis; gene therapy; protein therapy.
 XX
 OS Homo sapiens.
 XX
 XX US6331413-B1.
 PN 18-DEC-2001.
 PD 17-MAR-2000; 2000US-0527345.
 XX 17-MAR-1999; 99US-124820P.
 PR (ZYMO) ZYMOGENETICS INC.
 XX
 PI Adler DA, Sheppard PO;
 XX MPI: 2002-096707/13.
 DR N-PSDB; AHS20591.
 DR Polynucleotides encoding salivary proteins useful as anti-microbial
 PT agents -
 PT
 XX
 PS Claim 1; Column 49-52; 29pp; English.
 XX
 CC The invention relates to a polynucleotide derived from the 4q12-4q13
 CC region of human chromosome 4 and encoding a zsig63 polypeptide, a
 CC secreted salivary protein with anti-microbial activity. Due to their
 CC microbial activity, the sequences can be used in the study of microbial
 CC infections, e.g. for recombinant production of anti-microbial proteins.
 CC The sequences can be used in the treatment of tooth decay, periodontal
 CC disease, thrush, gastrointestinal disease, urinary tract infections,
 CC vaginal infections, skin infections, epithelial wounds, chronic tissue
 CC damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung
 CC infections, sarcoidosis, emphysema and chronic bronchitis. This sequence
 CC represents human zsig63.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

DR N-PSDB; ABX93594.
 XX Novel isolated zsig63 polypeptide, member of the adhesin family, useful
 PT for treating dental caries, periodontal disease, thrush,
 PT gastrointestinal disease, urinary tract infections, vaginal infections,
 PT skin infections
 XX
 XX Claim 10; Page 27-28; 32pp; English.
 XX
 CC The invention relates to an isolated zsig63 polypeptide comprising at
 CC least 90% identity to an amino acid sequence which comprises domain 1 of
 CC zsig63, domain 2, domain 3, mature zsig63 and full length zsig63. Also
 CC included are the polynucleotide encoding zsig63, a zsig63 expression
 CC vector, a cultured cell comprising the vector and expressing the protein,
 CC a DNA encoding a fusion protein (comprising amino acids 1-15, 16-37,
 CC 38-126, 127-219 or 16-219 of zsig63 and an additional protein), using a
 CC zsig63 reporter gene construct to identify zsig63 agonists, and
 CC producing an anti-zsig63 antibody using zsig63 immunogenic peptides,
 CC zsig63 is useful for detecting in a test sample, the presence of
 CC antagonist of zsig63 protein activity. Zsig63 has antimicrobial
 CC activity and since exhibits high expression in salivary gland, can be
 CC used for treating dental caries, periodontal disease, thrush, and
 CC gastrointestinal disease, urinary tract infections, vaginal infections,
 CC skin infections and other epithelial wounds. The polypeptides can be
 CC used to establish normal microflora and protect against pathogenic
 CC colonization and invasion. Zsig63 can also be used for providing
 CC pro-inflammatory activity for treating chronic, tissue damage
 CC particularly in areas having limited or damaged vascular system, e.g.
 CC in diabetes, and for treating immunocompromised AIDS patients or in
 CC individuals that have undergone chemotherapy, radiation treatment, for
 CC treating lung infections e.g. in cystic fibrosis. Detection of zsig63
 CC polypeptide at relatively high levels in the trachea may indicate that
 CC such polypeptides may serve as a marker of lung dysfunction. Zsig63 is
 CC also useful in diagnosing conditions associated with salivary gland or
 CC lung dysfunction including salivary gland carcinoma, pneumocystis carinii
 CC infection, emphysema, chronic bronchitis, prostate dysfunction such
 CC as prostate adenocarcinoma, aiding digestion, and as components of
 CC defined cell culture media and may be used to replace serum that is
 CC commonly used in culture. The DNA is useful in gene therapy applications
 CC to increase or inhibit zsig63 activity, and for detecting abnormalities
 CC on human chromosome 4 (e.g. 4q12-4q13, associated with dentinogenesis
 CC imperfecta, and dentin dysplasia type II). Zsig63 is an adhesin family
 CC member. The present sequence represents human zsig63.
 CC
 XX
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 24; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLWACTVCAFAARRRPPFGEDNDGHPHLSLNPYGIKRLPPLYYRPNVTP 60
 DB 1 MKLLWACTVCAFAARRRPPFGEDNDGHPHLSLNPYGIKRLPPLYYRPNVTP 60
 QY 61 SYGNTYTTDGLPSYWLITSPGFPYVYHIRGFLATQLVNPLPRGFPVPVPSRFFSA 120
 DB 61 SYGNTYTTDGLPSYWLITSPGFPYVYHIRGFLATQLVNPLPRGFPVPVPSRFFSA 120
 QY 121 AAAPAPPTAAAPAAAPPLTATVAAEPAAGAVAAEPAEAVGAEPAEAPVAAEPAA 180
 DB 121 AAAPAPPTAAAPAAAPPLTATVAAEPAAGAVAAEPAEAVGAEPAEAPVAAEPAA 180
 QY 181 EAPVGEPAEAPSPAPATAPKAPAPDEPHSPSLDAQNQ 219
 DB 181 EAPVGEPAEAPSPAPATAPKAPAPDEPHSPSLDAQNQ 219
 RESULT 6
 ID AAY19472 standard; Protein; 221 AA.
 XX
 AC AAY19472;
 XX

DT 14-JUL-1999 (first entry)
 XX
 DB Amino acid sequence of a human secreted protein.
 XX
 KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease;
 KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
 KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
 KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
 KW lung disease; thymus disease; digestive disorder; endocrine disorder;
 KW infection; AIDS.
 XX
 XX Homo sapiens.
 OS
 PN MO9922243-X1.
 XX
 PD 06-MAY-1999.
 XX
 XX 23-OCT-1998; 98WO-US22376.
 XX
 XX 24-OCT-1997; 97US-0063387.
 PR 24-OCT-1997; 97US-0062784.
 PR 24-OCT-1997; 97US-0063088.
 PR 24-OCT-1997; 97US-0063089.
 PR 24-OCT-1997; 97US-0063090.
 PR 24-OCT-1997; 97US-0063091.
 PR 24-OCT-1997; 97US-0063092.
 PR 24-OCT-1997; 97US-0063097.
 PR 24-OCT-1997; 97US-0063098.
 PR 24-OCT-1997; 97US-0063100.
 PR 24-OCT-1997; 97US-0063101.
 PR 24-OCT-1997; 97US-0063109.
 PR 24-OCT-1997; 97US-0063110.
 PR 24-OCT-1997; 97US-0063111.
 PR 24-OCT-1997; 97US-0063148.
 PR 24-OCT-1997; 97US-0063386.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Carter KC, Duan DR, Ebner R, Entress GA,
 PI Feng P, Florence C, Florence KA, Greene JM, Janat F,
 PI Kay H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA,
 PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
 XX
 DR WPI: 1999-303069/25.
 DR N-PSDB; AAX61352.
 XX
 XX New isolated human genes and the secreted polypeptides they encode
 PT
 PT Claim 11; Page 401-402; 546pp; English.
 PS
 CC The specification describes cDNA sequences (AAX61322-X61470) encoding
 CC human secreted proteins (AAY19442-Y19590). The polynucleotides and their
 CC corresponding secreted polypeptides are useful for preventing, treating
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.
 CC Pathological conditions can also be diagnosed by determining the amount
 CC of the polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the polynucleotides, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and fetal deficiencies, blood disorders, leukemias,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
 CC disorders involving osteoclasts such as osteoporosis, arthritis or
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
 CC disorders, infections and AIDS. The polypeptides are also useful for
 CC identifying their binding partners.
 CC
 XX
 SQ Sequence 221 AA;
 XX

Query Match 68.5%; Score 150; DB 20; Length 221;
Best Local Similarity 100.0%; Pred. No. 7.4e-135;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMAGTVCVAFARRRPFGEDDNDGHPHPSLNTIPYGRNLPPPLYRPNVTVP 60
DB 1 MKLLMAGTVCVAFARRRPFGEDDNDGHPHPSLNTIPYGRNLPPPLYRPNVTVP 60
QY 61 SYPGNTYDTGTPSYPMILTSRGPYYVYHIRGFLATQLVNPPPPRGFPVPSRPFSA 120
DB 61 SYPGNTYDTGTPSYPMILTSRGPYYVYHIRGFLATQLVNPPPPRGFPVPSRPFSA 120
QY 121 AAAPAPPIAAEPAAAPLTPVAAPAA 150
DB 121 AAAPAPPIAAEPAAAPLTPVAAPAA 150

RESULT 7
AAW30653
ID AAW30653 standard; Protein; 219 AA.
AC AAW30653;
DT 12-APR-1999 (first entry)
DE Human secreted protein clone cp16 1 protein.
KW Human; secreted protein; nutritional activity; cytokine; vaccine;
KW cell proliferation; differentiation; immune stimulation; suppression;
KW haemopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
KW chemokine; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
KW tumour invasion suppression; tumour inhibition.
OS Homo sapiens.
XX MO9901466-A1.
XX 14-JAN-1999.
XX 01-JUL-1998; 98MO-US13813.
XX 27-OCT-1997; 97US-0958304.
XX 02-JUL-1997; 97US-0887195.
PA (GENY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
PI Spaulding V, Treacy M;
DR WPI; 1999-105994/09.
DR N-PSDB; AAV80740.
XX New polynucleotides encoding secreted human proteins - are derived
XX from human foetal brain, adult testes, adult brain, foetal kidney,
XX adult salivary gland, or adult blood cDNA libraries, useful as, e.g.,
XX potential vaccines
XX
XX Claim 24; Page 71-72; 107pp; English.

XX The present sequence represents a human secreted protein from clone
XX cp16 1, deposited as ATCC 98482. Human secreted protein clone
XX polynucleotides and proteins are predicted to have biological
XX activities which would make them suitable for treating, preventing or
XX ameliorating medical conditions in humans and animals. Suggested
XX activities include nutritional activity, cytokine and cell
XX proliferation/differentiation activity, immune stimulating (e.g. as
XX vaccines) or suppressing activity, haematopoiesis regulating activity,
XX tissue growth activity, activin/inhibin activity, chemotactic/
XX chemokine activity, haemostatic and thrombolytic activity, receptor/
XX ligand activity, anti-inflammatory activity, cadherin/tumour invasion
XX suppressor activity, and tumour inhibition activity. The polynucleotides
XX are also stated to be useful for gene therapy.

XX SQ Sequence 219 AA;
Query Match 63.9%; Score 140; DB 20; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.5e-125;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMAGTVCVAFARRRPFGEDDNDGHPHPSLNTIPYGRNLPPPLYRPNVTVP 60
DB 1 MKLLMAGTVCVAFARRRPFGEDDNDGHPHPSLNTIPYGRNLPPPLYRPNVTVP 60
QY 61 SYPGNTYDTGTPSYPMILTSRGPYYVYHIRGFLATQLVNPPPPRGFPVPSRPFSA 120
DB 61 SYPGNTYDTGTPSYPMILTSRGPYYVYHIRGFLATQLVNPPPPRGFPVPSRPFSA 120
QY 121 AAAPAPPIAAEPAAAPLTPVAAPAA 140
DB 121 AAAPAPPIAAEPAAAPLTPVAAPAA 140

RESULT 8
ABB64202
ID ABB64202 standard; Protein; 75 AA.
AC ABB64202;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 19398.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX MO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL08305.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 19398; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins
XX (AB157737-AB172072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 75 AA;
Query Match 4.1%; Score 9; DB 22; Length 75;
Best Local Similarity 100.0%; Pred. No. 0.91;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
 DB 45 AAAAPAAP 53

RESULT 9
 AAM03688
 ID AAM03688 standard; Protein; 322 AA.
 AC AAM03688;
 XX
 AC AAM03688;
 XX
 DT 09-MAR-1997 (first entry)
 XX
 DE Leishmania chagasi acidic ribosomal antigen LcP0.
 XX
 KW Leishmania chagasi; acidic ribosomal antigen; LcP0;
 XX epitope; K39.
 XX
 OS Leishmania chagasi.
 XX

Key Location/Qualifiers
 FT CDS 30..1202
 FT /*tag= a

MO963414-A2.
 XX
 PD 24-OCT-1996.
 XX
 PF 19-APR-1996; 96MO-US05472.
 XX
 PR 21-APR-1995; 95US-0428414.
 XX
 PA (COR1-) CORIXA CORP.
 XX
 PI Reed SG;
 XX
 DR WPI; 1996-485884/48.
 XX
 DR N-PSDB; AAT42164.
 XX

PT New Leishmania acidic ribosomal P-protein family polypeptide - used
 to develop prods. for diagnosis, detection and protection against
 PT Leishmania infections
 XX
 PS Disclosure; Page 29-32; 76pp; English.
 XX

Compounds including polypeptides that contain at least an epitope of
 CC the L. chagasi acidic ribosomal antigen LcP0 are useful in a variety
 CC of immunoassays for detecting Leishmania infection. Portions of
 CC LcP0 (AAT42164) contg. at least the 17 C-terminal amino acids (AAT42165)
 CC have been found to generate a signal in an ELISA that is equivalent
 CC to that generated by the full length LcP0. A combination
 CC polypeptide may also be used, comprising an LcP0 epitope along with
 CC an epitope derived from the Leishmania K39 antigen (AAT42166), pref.
 CC the K39 repeat unit antigen having the sequence given in AAM03690.
 XX

SQ Sequence 322 AA;

Query Match 4.1%; Score 9; DB 17; Length 322;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAP 138
 DB 289 AAEPAAAP 297

RESULT 10
 ABB65430
 ID ABB65430 standard; Protein; 728 AA.
 XX
 AC ABB65430;

XX 26-MAR-2002 (first entry)
 DT
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 23082.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 PD 23-MAR-2001; 2001WO-US09231.
 XX
 PF 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PERK) PE CORP NY.
 PA
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR
 DR N-PSDB; ABL09533.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 23082; 21pp + Sequence Listing; English.
 XX

The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (AB01840-AB146175) and the encoded proteins
 CC (AB857737-AB872072).
 CC

The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC

SQ Sequence 728 AA;

Query Match 4.1%; Score 9; DB 22; Length 728;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAAP 127
 DB 418 SAAAPAAP 426

RESULT 11
 AAB76532
 ID AAB76532 standard; Protein; 1874 AA.
 XX
 AC AAB76532;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MCT protein SEQ ID NO:46.
 XX
 DB Corynebacterium glutamicum; brevbacterium; lacticofementum; MCT;
 XX membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering.
 XX
 OS Corynebacterium glutamicum.
 XX

XX AC ABP96450;
 XX DT 27-MAY-2003 (first entry)
 XX DE HLA class I molecule heavy chain optimal peptide #1.
 XX DE HLA class I molecule heavy chain optimal peptide #1.
 XX KM Human leukocyte antigen; HLA; major histocompatibility complex; MHC;
 XX KM T cell.
 XX OS Homo sapiens.
 XX OS Synthetic.
 XX PN WO2003016905-A2.
 XX PD 27-FEB-2003.
 XX PF 15-AUG-2002; 2002WO-GE03753.
 XX PR 16-AUG-2001; 2001GB-0020042.
 XX PA (AVID-) AVIDEX LTD.
 XX PI Jakobsen BK;
 XX DR WPI; 2003-268366/26.
 XX PT Determining whether a T-cell reacts with a predetermined Major
 PT Histocompatibility Complex (MHC) type, comprises contacting a sample of
 PT T-cell with MHC type molecules and determining whether the MHC
 PT molecules activate the T-cell -
 XX PS Disclosure; Page 17; 56pp; English.
 XX SS
 CC The present invention describes a method for determining whether a T-cell
 CC reacts with a predetermined major histocompatibility complex (MHC) type,
 CC which comprises bringing a sample comprising the T-cell into contact with
 CC several molecules of the MHC type, each MHC molecule being complexed with
 CC a peptide antigen whose contribution to a T-cell receptor binding the
 CC MHC-peptide antigen complex is minimised, and determining whether the MHC
 CC molecules cause activation of the T-cell. Also described: (1) a cloning
 CC vector encoding an MHC subunit (preferably human) into which a nucleotide
 CC encoding a peptide antigen sequence of interest can be inserted such that
 CC expression of the vector produces a fusion protein comprising the MHC
 CC subunit with the peptide antigen fused into it via a linker sequence;
 CC (2) a cell transformed with the above vector; (3) a multivalent class I
 CC or II MHC-peptide complex, preferably a multimer, in which the peptide
 CC antigen is not recognised in a specific manner by T-cells; and (4) a kit
 CC for carrying out the above method, comprising the plurality of MHC
 CC molecules cited above. The method is useful in determining whether a
 CC T-cell reacts with a predetermined major histocompatibility complex type,
 CC and in assessing allo-specific T-cell activity, or in matching transplant
 CC and donor patients, as well as in monitoring alloreactive responses
 CC following a transplant operation. The present sequence represents a human
 CC leukocyte antigen (HLA) class I molecule heavy chain optimal peptide,
 CC which is given in the exemplification of the present invention.
 XX SO Sequence 9 AA;
 XX
 Query Match 3.7%; Score 8; DB 24; Length 9;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 AAEPAAAA 137
 DB 1 AAEPAAAA 8
 RESULT 14
 ABP80080
 ID ABP80080 standard; Protein; 59 AA.
 XX
 AC ABP80080;

XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 6690.
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 6690.
 XX KW Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-1B02069.
 XX PR 12-FEB-2001; 2001GB-0003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Piazza M, Mastignani V, Monaci E;
 XX DR WPI; 2003-058415/05.
 XX DR N-PSDB; ABZ41050.
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX PT medicament for treating or preventing N. gonorrhoeae infection -
 XX PS Disclosure; Page 674; 815pp; English.
 XX SS
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX SO Sequence 59 AA;
 XX
 Query Match 3.7%; Score 8; DB 24; Length 59;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 198 PATAKPAA 205
 DB 25 PATAKPAA 32
 RESULT 15
 AAG27598
 ID AAG27598 standard; Protein; 65 AA.
 XX
 AC AAG27598;
 XX DT 17-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 32499.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX PR 25-FEB-1999; 99US-0121825.
 XX PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128233.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134761.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142053.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143547.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147250.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147943.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155119.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161892.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.7%; Score 8; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 AAEPAADA 137
Db 19 AAEPAADA 26

Search completed: January 15, 2004, 15:28:31
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:23:11 ; Search time 18 Seconds

(without alignments)
572.158 Million cell updates/sec

Title: US-09-923-236-2

Sequence: 1 MCLLMACIVCAVAFARKRF.....TAKPAPEHPSPSLQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	9	4.1	RLA0_LEICH	P39086 leishmania
2	9	4.1	RLA0_LEICH	P39097 leishmania
3	9	4.1	HEMI_STRCO	Q6wx15 streptomyce
4	9	4.1	NIA_VOLCA	P6841 volvox cart
5	8	3.7	RLA1_LEIPE	Q66313 leishmania
6	8	3.7	RLA3_SCHPO	P17477 schistosch
7	8	3.7	RL7_CHLBN	Q32961 chlamydia p
8	8	3.7	CUP5_GALMB	Q24988 galliera m
9	8	3.7	P53_MESAV	Q00366 mesocricetu
10	8	3.7	ATP2_CHIRB	P38482 chlamydomon
11	8	3.7	MSPA_TRLMA	Q92413 treponema m
12	8	3.7	BCR_HELYT	Q18473 heliothis v
13	8	3.7	KELC_DROME	Q04652 drosophila
14	8	3.7	NTC2_MOUSE	Q35516 mus musculu
15	8	3.2	ANPA_PSBAM	P04002 pseudopleur
16	7	3.2	RLI2_MERTH	P05394 methanobact
17	7	3.2	RLI2_MERTL	Q52706 methanococ
18	7	3.2	RLI2_MERTL	P54048 methanococ
19	7	3.2	RLA2_DICDI	P22683 dictyostella
20	7	3.2	RLA1_CHLRE	P29763 chlamydomon
21	7	3.2	RLA1_TRYCR	P66643 trypanosoma
22	7	3.2	RLA2_CRYST	O61463 cryptochito
23	7	3.2	RAZA_MAIZE	P46252 zea mays (m
24	7	3.2	RLI2_ABRPE	Q93949 aeropyrum p
25	7	3.2	RLI1_CABEL	P19193 caenorhabdi
26	7	3.2	RLA2_ASPFU	Q9406 aspergillus
27	7	3.2	RLA1_DROME	P08570 drosophila
28	7	3.2	RLI2_HALNI	P05768 halobacteri
29	7	3.2	RLA1_RAT	P19944 rattus norv
30	7	3.2	RLA2_BIMTE	Q96749 eimeria ten
31	7	3.2	CU19_ARADI	P80515 araneus dia
32	7	3.2	PSAE_SPTOL	P12354 epinacia ol
33	7	3.2	DYLA_CHLRE	Q93591 chlamydomon

34	7	3.2	136	1	RL28 HUMAN	P46779 homo sapien
35	7	3.2	136	1	RL28 MOUSE	P41105 mus musculu
36	7	3.2	136	1	SL14_HUMAN	P37108 homo sapien
37	7	3.2	156	1	BCCP_PSEAE	P37789 pseudomonas
38	7	3.2	163	1	ATPD_CABEL	Q09544 caenorhabdi
39	7	3.2	168	1	ATPD_HUMAN	P30049 homo sapien
40	7	3.2	199	1	CYCY_RHOCA	Q05389 rhodobacter
41	7	3.2	204	1	RS3A_METWA	Q89052 methanosarc
42	7	3.2	208	1	RS6_MYCGB	P47336 mycoplasma
43	7	3.2	215	1	ERPD_WHEAT	P29546 triticum ae
44	7	3.2	215	1	RS6_MYCPN	P75543 mycoplasma
45	7	3.2	220	1	COAT_CMVSI	Q00467 cymbidium m

ALIGNMENTS

```

RESULT 1
RLA0_LEICH          STANDARD;   PRT;   322 AA.
AC P39096;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 60S acidic ribosomal protein P0.
GN LCPO.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE MHOW/BR/82/BA-2;
RX MEDLINE=94222525; PubMed=7513304;
RA Skelly Y.A.W., Benson D.R., Elvasilla M., Badaro R., Burns J.M. Jr.,
RA Reed S.G.;
RT "Antigens shared by leishmania species and Trypanosoma cruzi:
RT Immunological comparison of the acidic ribosomal P0 proteins.";
RL Infect. Immun. 62:1643-1651(1994).
CC -!- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF R. COLI PROTEIN L10.
CC -!- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
CC DIMERS OF P1 AND P2.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L29300; AAA29262.1; -
DR InterPro: IPR001813; 60S_Ribosomal.
DR InterPro: IPR001790; Ribosomal_L10.
DR Pfam: PF00428; 60S_Ribosomal; 1.
DR Pfam: PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Phosphorylation.
SQ
SEQUENCE 322 AA; 34594 MW; 2389P125356D26D2 CRC64;

Query Match          4.1%; Score 9; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAP 138
   |||||
Db 289 AAEPAAP 297

RESULT 2
RLA0_LEICH          STANDARD;   PRT;   323 AA.
AC P39097;

```

DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 60S acidic ribosomal protein P0.
 GN L1PO-A AND L1PO-B.
 OS Leishmania infantum.
 OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=5671;
 RX MEDLINE=94088674; PubMed=8264730;
 RA Soto M., Requena J.M., Alonso C.;
 RT "Isolation, characterization and analysis of the expression of the
 RL Leishmania ribosomal P0 protein genes.";
 M1. Biochem. Parasitol. 61:265-274(1993).
 CC -1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
 CC OF E. COLI PROTEIN L10.
 CC -1- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
 CC DIMERS OF P1 AND P2.
 CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X72714; CAA51264.1; -;
 CC EMBL; X72714; CAA51264.1; -;
 CC InterPro; IPR001813; 60S ribosomal.
 CC InterPro; IPR001790; Ribosomal L10.
 CC Pfam; PF00428; 60S_rribosomal; 1.
 CC Pfam; PF00466; Ribosomal_L10; 1.
 CC Ribosomal protein; Phosphorylation.
 KW RIBOSOMAL PROTEIN; PHOSPHORYLATION.
 SQ SEQUENCE 323 AA; 34771 MW; 14967BD7A439D69E CRC64;
 Query Match 4.1%; Score 9; DB 1; Length 323;
 Best local similarity 100.0%; Pred. No. 0.66; Mismatches 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 AAEPAAAP 138
 DB 290 AAEPAAAP 298
 RESULT 3
 HEMI_STRCO STANDARD; PRT; 581 AA.
 ID HEMI_STRCO
 AC Q9WX15;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GLUTR).
 GN HEMA OR SC03319 OR SCE68.17C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 NCBI_TaxID=1902;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch B., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Batteil B.G., Parkhill J.,

RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
 CC semialdehyde + NADP(+) + tRNA(Glu).
 CC -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
 CC -1- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL939116; CAB45353.1; -;
 CC FIR; T36267; T36267.
 CC HAMAP; MF_00087; atypical; 1.
 CC InterPro; IPR000343; GlutR.
 CC InterPro; IPR000594; Thif domain.
 CC Pfam; PF00745; GlutR_dimer; 1.
 CC Pfam; PF05201; GlutR_N; 1.
 CC Pfam; PF05200; GlutR_NAD_bind; 1.
 CC TIGRPFAM; TIGR01035; hemaA; 1.
 CC PROSITE; PS00747; GLUTR; 1.
 KW Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT DOMAIN 292 416 INSERT.
 FT ACT SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 99 99 BASE (BY SIMILARITY).
 SQ SEQUENCE 581 AA; 60562 MW; D48256B105AFA037 CRC64;
 Query Match 4.1%; Score 9; DB 1; Length 581;
 Best local similarity 100.0%; Pred. No. 1.1; Mismatches 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 154 VAAEPAAEA 162
 DB 36 VAAEPAAEA 44
 RESULT 4
 NIA_VOLCA STANDARD; PRT; 864 AA.
 ID NIA_VOLCA
 AC P36841;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nitrate reductase (NADH) (EC 1.7.1.1) (NR).
 GN NITA.
 OS Nitrospira.
 OC Bacteria; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Volvocaceae; Volvox.
 NCBI_TaxID=3067;
 RX MEDLINE=93013022; PubMed=1398126;
 RA Gruber H., Goetlink S.D., Kirk D.L., Schmitt R.;
 RT "The nitrate reductase-encoding gene of Volvox carterii: map location,
 RT sequence and induction kinetics.";
 RL Gene 120:75-83(1992).
 CC -1- FUNCTION: Nitrate reductase is a key enzyme involved in the first
 CC step of nitrate assimilation in plants, fungi and bacteria.
 CC -1- CATALYTIC ACTIVITY: Nitrite + NAD(+) + H(2)O = nitrate + NADH.
 CC HEM IRON, AND POLYDENDROM-PTERIN AS PROSTHETIC GROUPS. THE HEME
 CC GROUP IS CALLED CYTOCHROME B-557.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- INDICATION: By nitrate.
 CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOTERIN OXIDOREDUCTASES IN THE
 CC N-TERMINAL DOMAIN.

```

CC -1- SIMILARITY: Contains 1 cytochrome b5 heme-binding domain.
CC -1- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; X64136; CA45497.1; -.
DR PIR; JCI422; JCI422.
DR HSSP; P04166; 1BSM.
DR InterPro; IPR001199; Cyt_B5.
DR InterPro; IPR001834; Cyt_B5_reductase.
DR InterPro; IPR000572; Euk_Mb_oxred.
DR InterPro; IPR001709; FPN_Cyt_reductase.
DR InterPro; IPR005066; Mo-co_dimer.
DR InterPro; IPR001433; Oxred_FAD/NAD(P).
DR Pfam; PF00970; FAD_binding_6; 1.
DR Pfam; PF00173; heme_1; 1.
DR Pfam; PF03404; Mo-co_dimer; 1.
DR Pfam; PF00175; NAD_binding_1; 1.
DR Pfam; PF00174; oxidored_molyb; 1.
DR PRINTS; PR00406; CYTB5REDTASE.
DR PRINTS; PR00363; CYTOCHROMEBS.
DR PRINTS; PR00407; EMOPTERIN.
DR PRINTS; PR00371; FPNCR.
DR ProDom; PD000612; Cyt_B5; 1.
DR PROSITE; PS00191; CYTOCHROME_B5_1; 1.
DR PROSITE; PS00255; CYTOCHROME_B5_2; 1.
DR PROSITE; PS00559; MOLYBDOTERIN_EUK; 1.
DR Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
DR Nitrate assimilation.
FT METAL 139 139 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 193 193 MOLYBDENUM-PTERIN (POTENTIAL).
FT DISULFID 376 376 INTERCHAIN (POTENTIAL).
FT METAL 532 532 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT METAL 555 555 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 864 AA; 96402 MW; 499529652CDD1C7 CRC64;

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
Db 478 AAAAPAAP 486

RESULT 5
RLAL_LEIPE STANDARD; PRT; 107 AA.
AC 046313;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S acidic ribosomal protein p1.
OS Leishmania peruviana.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOM/PE/84/LC26;
RA de los Santos M., Carrillo C., Panebra A., Montoya Y.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; AF045249; AAC02701.1; -.
DR InterPro; IPR001813; 60S_ribosomal.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00428; 60S_ribosomal; 1.
DR PRINTS; PR00456; RIBOSOMALP2.
DR Ribosomal protein.
SQ SEQUENCE 107 AA; 10896 MW; 9117DEC5E490A071 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAPA 126
Db 71 SAAAPAPA 78

RESULT 6
RLA3_SCHPO STANDARD; PRT; 110 AA.
AC P17477;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein p1-alpha 3 (A3).
GN RPA3 OR SPBC3B9.13C.
OS Schizosaccharomyces pombe (Pission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220620; PubMed=2325655;
RA Beltrame M., Bianchi M.B.;
RT "A gene family for acidic ribosomal proteins in Schizosaccharomycetes
RT pombe: two essential and two nonessential genes.";
RL Mol. Cell. Biol. 10:2341-2348(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream W.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle R.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moute S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckert G., Aert R., Robben J., Gymnopoulos B.,
RA Weitzens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mosati D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Reger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Kochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

```

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schistosoma mansoni pomb.";
 CC Nature 415:871-880(2002).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
 CC PROTEIN SYNTHESIS.
 CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
 CC SUBUNIT.
 CC -1- MISCELLANEOUS: YEASTS CONTAIN A INDIVIDUAL SMALL RIBOSOMAL A
 CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO
 CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.
 CC -1- MISCELLANEOUS: RPA3 AND RPA4 ARE ESSENTIAL FOR CELL SURVIVAL.
 CC WHEREAS RPA1 AND RPA2 ARE NOT.
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M31319; AAA35336.1; -;
 CC EMBL; AL022070; CA117793.1; -;
 CC PIR; C34715; R6BYP3.
 CC GeneDB_Sprobe; SPBC3B9.13c; -;
 CC InterPro; IPR001813; 609_Ribosomal.
 CC Pfam; PF00428; 608_Ribosomal.1.
 CC KMW Ribosomal protein; Phosphorylation; Multigene family.
 CC SQ SEQUENCE 110 AA; 11171 MW; 1734AC9779F5A891 CRC64;
 CC
 CC Query Match 3.7%; Score 8; DB 1; Length 110;
 CC Best Local Similarity 100.0%; Pred. No. 2.3;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 119 SAAAPPA 126
 CC Db 64 SAAAPPA 71
 CC
 CC RESULT 7
 CC ID RL7 CHLPN STANDARD; PRT; 128 AA.
 CC AC 092361; 09J070;
 CC DT 30-MAY-2000 (Rel. 39, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE 50S ribosomal protein L7/L12.
 CC GN RPL7 OR RL7 OR CPN080 OR CP0695.
 CC OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 CC OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC NCBI_TaxID=83558;
 CC RX MEDLINE=99206606; PubMed=10192388;
 CC RA Kaimen S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 CC RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 CC RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 CC RL Nat. Genet. 21:385-389(1999).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC RP STRAIN=AR39;
 CC RC MEDLINE=20150255; PubMed=10684935;
 CC RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 CC RA White O., Hickey E.K., Peterson J., Uetrich T., Berry K., Bass S.,
 CC RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 CC RA Linher M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 CC RA Bisen J., Fraser C.M.;
 CC RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
 CC pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=J138;
 RC MEDLINE=20330349; PubMed=10871362;
 RA Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWT029 from USA";
 CC Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
 CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
 CC ACCURATE TRANSLATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001593; AAD18233.1; -;
 CC EMBL; AE002228; AAF38503.1; -;
 CC EMBL; AP002545; BAA98290.1; -;
 CC PIR; C72122; C72122.
 CC PIR; H66500; H66500.
 CC HSRP; P02392; ICTP.
 CC PHC1-2DPAGE; Q929A1; -;
 CC TIGR; CP0695; -;
 CC HAMAP; MF_00368; -; 1.
 CC InterPro; IPR000206; Ribosomal_L12.
 CC Pfam; PF00542; Ribosomal_L12.
 CC ProDom; PD001326; Ribosomal_L12; 1.
 CC TIGRPAWS; TIGR00855; L12; 1.
 CC KMW Ribosomal protein; Complete proteome.
 CC FT INIT MET 0
 CC BY SIMILARITY.
 CC SQ SEQUENCE 126 AA; 13461 MW; 4B2F17A85B057CC CRC64;
 CC
 CC Query Match 3.7%; Score 8; DB 1; Length 126;
 CC Best Local Similarity 100.0%; Pred. No. 2.6;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 171 EAPVAAP 178
 CC Db 49 EAPVAAP 56
 CC
 CC RESULT 8
 CC ID CUPS_GALME STANDARD; PRT; 353 AA.
 CC AC 024958; 024958;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE Pupal cuticle protein PCP52 precursor (GMPCP52).
 CC GN PCP52.
 CC OS Galleria mellonella (Wax moth).
 CC OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyralidae;
 CC OC Pyralidae; Galleriinae; Galleria.
 CC NCBI_TaxID=7137;
 CC RX MEDLINE=95291282; PubMed=7773255;
 CC RA Kolberg U., Obermaier B., Hirsch H., Kelber G., Wolbert P.;
 CC RT "Expression cloning and characterization of a pupal cuticle protein
 CC cDNA of Galleria mellonella L.";
 CC RT Insect Biochem. Mol. Biol. 25:355-363(1995).
 CC -1- FUNCTION: COMPONENT OF THE CUTICLE OF THE PUPA OF GALLERIA
 CC MELLONELLA.

CC -1- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION DURING THE FIRST DAY AFTER
CC PIPAL ECYOSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77514; CAAS4650.1; -.
DR PIR; S41958; S41958.
KW Structural protein; Cuticle; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 353 PUPAL CUTICLE PROTEIN PCP52.
FT DOMAIN 235 243 POLY-ALA.
SQ SEQUENCE 353 AA; 35818 MW; C3A35B04EC049172 CRC64;
Query Match 3.7%; Score 8; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 119 SAAAPPA 126
Db 236 SAAAPPA 243
RESULT 9
P53_MESAU STANDARD; PRT; 396 AA.
AC Q00366; P97276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellular tumor antigen p53 (Tumor suppressor p53).
GN TP53.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euhetia; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SYRIAN; TISSUE=Kidney;
RX MEDLINE=92210007; Pubmed=1555773;
RA Legros Y., McIntyre P., Soussi T.;
RL "The CDNA cloning and immunological characterization of hamster p53."
RN Gene 112:247-250(1992).
RP SEQUENCE FROM N.A.
RA Hou E.W., Wiseeman R.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Acts as a tumor suppressor in many tumor types; induces
CC growth arrest or apoptosis depending on the physiological
CC circumstances and cell type. Involved in cell cycle regulation as
CC a trans-activator that acts to negatively regulate cell division
CC by controlling a set of genes required for this process. One of
CC the activated genes is an inhibitor of cyclin-dependent kinases.
CC Apoptosis induction seems to be mediated either by stimulation of
CC BAX and P53 antigen expression, or by repression of Bcl-2
CC expression.
CC -1- SUBUNIT: Binds DNA as a homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- DISBURS: p53 is found in increased amounts in a wide variety
CC of transformed cells. p53 is frequently mutated or inactivated
CC in many types of cancer.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75144; AAA37085.1; -.
DR EMBL; U07182; AAB41344.1; -.
DR PIR; JH0633; JH0633.
DR HSSP; P04637; ITUP.
DR InterPro; IPR02117; P53.
DR Pfam; PF00870; P53; 1.
DR PRINTS; PR00386; P53SUPPRESSR.
DR ProDom; PD002681; P53; 1.
DR PROSITE; PS00348; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 45 TRANSCRIPTION ACTIVATION (ACIDIC).
FT DNA_BIND 105 295 BY SIMILARITY.
FT DOMAIN 328 359 OLIGOMERIZATION.
FT DOMAIN 371 390 BASIC (REPRESSION OF DNA-BINDING).
FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 15 15 PHOSPHORYLATION (BY PRPK) (BY
FT MOD_RES 395 395 SIMILARITY).
FT MOD_RES 188 188 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 188 188 G -> S (IN REF. 2).
SQ SEQUENCE 396 AA; 43631 MW; 906EF02568099BE3 CRC64;
Query Match 3.7%; Score 8; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 120 AAAAPAP 127
Db 66 AAAAPAP 73
RESULT 10
ATP2_CHLRE STANDARD; PRT; 574 AA.
AC P384B2;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14).
GN ATP2.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92353366; Pubmed=1386535;
RA Franzen L.-G., Falk G.;
RL "Nucleotide sequence of cDNA clones encoding the beta subunit of
RL mitochondrial ATP synthase from the green alga Chlamydomonas
RL reinhardtii: the precursor protein encoded by the cDNA contains both
RL an N-terminal presequence and a C-terminal extension.";
RN Plant Mol. Biol. 19:771-780(1992).
RP STRUCTURE BY NMR OF 1-26.
RX MEDLINE=96326639; Pubmed=8706917;
RA Lancelin J.-M., Gans P., Bouchayer E., Bally I., Arlaud G.J.,
RA Jacquot J.-P.;
RL "NMR structures of a mitochondrial transit peptide from the green
RL alga Chlamydomonas reinhardtii.";
RL PDBS Lett. 391:203-208(1996).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)

```

CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC      -1- SUBCELLULAR LOCATION: Mitochondrial.
CC      -1- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X61624; CAA43808.1; -.
CC      DR PIR; S23530; S23530.
CC      DR HSSP; P00829; IBMF.
CC      DR InterPro; IPR003593; AAA_ATPase.
CC      DR InterPro; IPR000793; ATPase_a/bc.
CC      DR InterPro; IPR000194; ATPase_a/bc.
CC      DR InterPro; IPR004100; ATPase_a/bn.
CC      DR InterPro; IPR005722; F1_ATPase_beta.
CC      DR Pfam; PF00006; ATP-synt_ab_C; 1.
CC      DR Pfam; PF00306; ATP-synt_ab_C; 1.
CC      DR Pfam; PF02874; ATP-synt_ab_N; 1.
CC      DR SMART; SM00382; AAA; 1.
CC      DR TIGRfam; TIGR01039; atpD; 1.
CC      DR PROSITE; PS00152; ATPASE ALPHA BETA; 1.
CC      KM ATP synthesis; CF(1); Hydrogen ion transport;
CC      KW Hydrolyase; ATP-binding; Mitochondrion; Transient.
CC      FT TRANSIT 1 26 MITOCHONDRION (POTENTIAL).
CC      FT CHAIN 27 574 ATP SYNTHASE BETA CHAIN.
CC      FT NP BIND 183 190 ATP (BY SIMILARITY).
CC      SQ SEQUENCE 574 AA; 61821 MW; 2283C6C6D18FBCFE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 9.2; Length 574;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      130 AAEPAAAA 137
DB      26 AAEPAAAA 33

RESULT 11
MSPA TREMA STANDARD; PRT; 575 AA.
AC 092413;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Major outer membrane protein mspa precursor (Major sheath protein).
GN MSPA.
OS Treponema maltophilum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=51100;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 20-34.
RC STRAIN=ATCC 51939 / BR;
RX MEDLINE=99121045; PubMed=9922270;
RA Heuner K., Choi B.-K., Schade R., Møter A., Otto A., Goebel U.B.;
RT Cloning and characterization of a gene (mspa) encoding the major
RT sheath protein of Treponema maltophilum ATCC 51939 (T)";
RT J. Bacteriol. 181:1025-1029 (1999).
RN [2]
RN SUBCELLULAR LOCATION.
RX MEDLINE=21213882; PubMed=1131333;
RA Heuner K., Meltzer U., Choi B.-K., Goebel U.B.;
RT "Outer sheath associated proteins of the oral spirochete Treponema
RT maltophilum.";
RL FEMS Microbiol. Lett. 197:187-193 (2001).
CC -1- FUNCTION: Major component of the outer membrane sheath.
CC -1- SUBCELLULAR LOCATION: Outer membrane-associated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Y17800; CAA76862.1; -.
CC      DR EMBL; Y17800; CAA76862.1; -.
CC      FT SIGNAL 1 19
CC      FT CHAIN 20 575
CC      SQ SEQUENCE 575 AA; 62250 MW; C666B18875DE16A0 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 9.2; Length 575;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      155 AAEPAAEA 162
DB      19 AAEPAAEA 26

RESULT 12
ECR HELVI STANDARD; PRT; 576 AA.
AC 018473;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ecdysone receptor (Ecdy/steroid receptor) (20-hydroxy-ecdysone
DE receptor) (20R receptor) (HVECR).
GN ECR OR NR1H1.
OS Heliothis virescens (Noctuid moth) (Owl moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Heliothinae; Heliothis.
OX NCBI_TaxID=7102;
RN [1]
RN SEQUENCE FROM N.A.
RN STRAIN=BR;
RX MEDLINE=99457718; PubMed=10528411;
RA Martinez A., Scanlon D., Gross B., Petras S.C., Balli S.R.,
RA Greenland A.V., Windas J., Pong O., Broad P., Jepson I.;
RT "Transcriptional activation of the cloned Heliothis virescens
RT (Lepidoptera) ecdysone receptor (HVECR) by murtisteroneA.";
RT Insect Biochem. Mol. Biol. 29:915-930 (1999).
CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
CC ELEMENTS (ECREs) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR1
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC      EMBL; Y09009; CAA70212.1; -.
CC      DR HSSP; P20393; IAEY.
CC      DR InterPro; IPR000536; Hormone_rec_1lg.
CC      DR InterPro; IPR001723; Strhma_receptor.
CC      DR InterPro; IPR001628; znf_Casteroid.
CC      DR Pfam; PF00104; hormone_rec; 1.
CC      DR Pfam; PF00105; zf-C4; 1.
CC      DR PRINTS; PR00398; STRDHOMNER.
CC      DR PRINTS; PR00047; STROIDFINGER.
CC      DR Prodom; PD000035; znf_Casteroid; 1.
CC      DR SMART; SM00430; HOL1; 1.
CC      DR SMART; SM00399; znf_C4; 1.
CC      DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.

```


Kw Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KZ Zinc-finger.
 FM DOMAIN 1 162 MODULATING (POTENTIAL).
 FT DNA BIND 163 228 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 163 228 C4-TYPE.
 FT ZN_FING 199 223 C4-TYPE.
 FT DOMAIN 326 545 HORMONE-BINDING (POTENTIAL).
 SQ SEQUENCE 576 AA; 64638 MW; D3EE787BF263A8 CRC64;
 Query March 3.7%; Score 8; DB 1; Length 576;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 140 TATPVAE 147
 Db 552 TATPVAE 559
 RESULT 13
 ID KELC DROME STANDARD; PRT: 1477 AA.
 AC 004653; 004653; 09VUA2;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ring canal kelch protein [Contains: Kelch short protein].
 GN KEL OR CG7210.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN SEQUENCE FROM N.A.
 RC TISSUE=Embryo.
 RX MEDLINE=93201392; PubMed=8453663;
 RA Xue F., Cooley L.;
 RT "kelch encodes a component of intercellular bridges in Drosophila egg
 chambers."
 RL Cell 72:681-693(1993).
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Aamam M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anagnostis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballow R.M., Baer A., Baxendale J., Bayraktarglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brinkstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Canter A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelle S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Syrakas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassenaar D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [3]
 RP CHARACTERIZATION.
 RC TISSUE=Embryo;
 RX MEDLINE=97236487; PubMed=9118811;
 RA Robinson D.N., Cooley L.;
 RT "Examination of the function of two kelch proteins generated by stop
 codon suppression."
 RL Development 124:1405-1417(1997).
 CC -1- FUNCTION: COMPONENT OF RING CANALS THAT REGULATES THE FLOW OF
 CYTOPLASM BETWEEN CELLS. MAY BE INVOLVED IN THE REGULATION OF
 CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.
 CC BINDS ACTIN.
 CC -1- SUBCELLULAR LOCATION: INNER SURFACE OF CYTOPLASMIC BRIDGES OR RING
 CANALS PRESENT IN EGG CHAMBERS. SUBCORTICALLY IN IMAGINAL DISK
 EPITHELIA.
 CC -1- TISSUE SPECIFICITY: BOTH PROTEINS ARE EXPRESSED IN OVARIES, MALE
 TESTIS, OVARECTOMIZED FEMALES, CUTICLE, SALIVARY GLAND AND
 IMAGINAL DISKS. KELCH ORP1 IS THE PREDOMINANT PROTEIN AND IS ALSO
 EXPRESSED IN FAT BODIES. ON ENTRY INTO METAMORPHOSIS LEVELS OF
 FULL LENGTH PROTEIN INCREASE IN TESTIS AND IMAGINAL DISKS.
 CC -1- DEVELOPMENTAL STAGE: LARVAE, PUPAE AND ADULTS.
 CC -1- SIMILARITY: Contains 6 BTB/POZ domain.
 CC -1- CAUTION: REF.3 BELIEVES RESIDUE 690 IS A SELENOCYSTEINE.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 gene model prediction.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L08483; AAA53471.1; -
 CC EMBL: L08483; AAA53472.1; -
 CC EMBL: AE003657; AAF53651.1; ALT_SEQ.
 CC HSSP: Q05516; 1C63.
 CC PDBase: PDB0001301; Kel.
 CC GO: 0007292; P: oogenesis; IMP.
 CC GO: 0007301; P: ring canal formation; IDA.
 CC InterPro: IPR000210; BTB_POZ.
 CC InterPro: IPR006511; Kelch.
 CC InterPro: IPR006512; Kelch_rep.
 CC Pfam: PF00651; BTB; 1.
 CC Pfam: PF01344; Kelch; 6.
 CC PRINTS: PR00501; KELCHREPEAT.
 CC SMART: SM00225; BTB; 1.
 CC SMART: SM00612; Kelch; 6.
 CC PROSITE: PS50057; BTB; 1.
 CC Cytochrome; Actin-binding; Selenium; Selenocysteine; Kelch repeat;
 KW Repeat.
 KM
 FT CHAIN 1 1477 KELCH PROTEIN.
 FT CHAIN 1 689 KELCH SHORT PROTEIN.
 FT REPEAT 157 223 BTB.
 FT REPEAT 404 449 KELCH 1.
 FT REPEAT 450 496 KELCH 2.
 FT REPEAT 498 543 KELCH 3.
 FT REPEAT 545 592 KELCH 4.
 FT REPEAT 594 639 KELCH 5.
 FT REPEAT 641 687 KELCH 6.
 FT REPEAT 18 28 ASN-RICH.
 FT DOMAIN 29 87 GLN-RICH.

```

FT DOMAIN 29 36 POLY-GLN.
FT DOMAIN 78 83 POLY-GLN.
FT SE CYS 690 690 PROBABLE.
FT CONFLICT 493 493 V -> A (IN REF. 1).
FT CONFLICT 596 596 A -> R (IN REF. 1).
FT CONFLICT 824 824 P -> L (IN REF. 1).
FT CONFLICT 858 858 G -> D (IN REF. 1).
FT CONFLICT 1083 1083 A -> R (IN REF. 1).
FT CONFLICT 1086 1086 A -> G (IN REF. 1).
SQ SEQUENCE 1477 AA; 160086 MW; 4851BEAE9D9DBA7 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 1477;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 120 AAAAPAP 127
Db 740 AAAAPAP 747

RESULT 14
NTC2 MOUSE STANDARD; PRT; 2470 AA.
ID NTC2_MOUSE Q6C09A1;
AC Q35516; Q6C09A1; Q6C09A1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
DE B)
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene."
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RA MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch 1 and Notch 2 mouse Notch homologues coexpressed in a
RT wide variety of tissues."
RL Exp. Cell Res. 204:364-372 (1993).
RN [3]
RP SEQUENCE OF 1765-2153 FROM N.A.
RC MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1."
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019 (1996).
RN [4]
RP FUNCTION.
RC MEDLINE=99396706; PubMed=10393120;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT embryonic lethality."
RL Development 126:3415-3424 (1999).
RN [5]
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RC MEDLINE=95333893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
RT mouse brain."
RL Brain Res. Mol. Brain Res. 29:263-272 (1995).
RN [6]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

```

```

RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis."
RL J. Biol. Chem. 276:40268-40273 (2001).
RN [7]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RC MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(IEC) which are probably linked by disulfide
CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1;
CC IsoId=Q35516-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q35516-2; Sequence=VSP_001405;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC neuroepithelia, somites, optic vesicles and branchial arches, but
CC not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC the postnatal ependymal cells, and the choroid plexus throughout
CC embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TW) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXN). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: Contains 35 EGF-like domains.
CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
CC EMBL: D32210; BAA22094.1; -
CC EMBL: X68279; CA48340.1; -
CC EMBL: U31881; AAC52924.1; -
CC PIR: A49175; A49175.
CC HSP: P16109; 1FSB.
CC MGD: MGI:97364; Notch2.
CC GO: GO:0005887; C:integral to plasma membrane; IC.
CC GO: GO:0005515; F:protein binding activity; IP.
CC GO: GO:0002011; F:morphogenesis of an epithelial sheet; IMP.

```

DR GO; GO:0007219; P:N receptor signaling pathway; IC.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR02049; Laminin_EGF.
 DR InterPro; IPR00800; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBLDOD.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50086; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 22.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2470 NOTCH INTRACELLULAR TRUNCATION.
 FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
 FT TRANSLEM 26 1677 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-Like 1.
 FT DOMAIN 64 102 EGF-Like 2.
 FT DOMAIN 105 143 EGF-Like 3.
 FT DOMAIN 144 180 EGF-Like 4.
 FT DOMAIN 182 219 EGF-Like 5, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 221 256 EGF-Like 6 (INCOMPLETE).
 FT DOMAIN 258 294 EGF-Like 7, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 296 334 EGF-Like 8, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 336 372 EGF-Like 9, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 373 411 EGF-Like 10.
 FT DOMAIN 413 452 EGF-Like 11, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 454 490 EGF-Like 12, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 492 528 EGF-Like 13, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 530 566 EGF-Like 14, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 568 603 EGF-Like 15, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 605 641 EGF-Like 16, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 643 678 EGF-Like 17, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 680 716 EGF-Like 18, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 718 753 EGF-Like 19.
 FT DOMAIN 753 791 EGF-Like 20, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 793 829 EGF-Like 21, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 831 869 EGF-Like 22.
 FT DOMAIN 871 907 EGF-Like 23, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 909 945 EGF-Like 24, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 947 983 EGF-Like 25, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 985 1021 EGF-Like 26, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1023 1059 EGF-Like 27, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1061 1097 EGF-Like 28.
 FT DOMAIN 1099 1145 EGF-Like 29.
 FT DOMAIN 1147 1183 EGF-Like 30, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1185 1221 EGF-Like 31, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1223 1260 EGF-Like 32, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1262 1300 EGF-Like 33.
 FT DOMAIN 1302 1345 EGF-Like 34.
 FT DOMAIN 1372 1410 EGF-Like 35.
 FT REPEAT 1418 1454 LIN/NOTCH 1.
 FT REPEAT 1501 1533 LIN/NOTCH 2.
 FT REPEAT 1825 1869 ANK 1.

FT REPEAT 1874 1903 ANK 2.
 FT REPEAT 1907 1937 ANK 3.
 FT REPEAT 1941 1970 ANK 4.
 Query Match 3.7%; Score 8; DB 1; Length 2470;
 Best Local Similarity 100.0%; Pred. No.32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 120 AAAAPAP 127
 Db 2183 AAAAPAP 2190
 RESULT 15
 ANPA_PSEAM STANDARD; PRT; 82 AA.
 ID ANPA_PSEAM
 AC P04002;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antifreeze protein A/B precursor.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; OC Acanthomorpha; Acanthopterygii; Perciformes; Pleuronectiformes; OC Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
 OX NCBI_TaxId=8265;
 RN [1]
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=82197490; PubMed=6952188;
 RA Davies P.L., Roach A.H., Hew C.-L.;
 RT "DNA sequence coding for an antifreeze protein precursor from winter flounder."
 RL Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
 RN [2]
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=88259236; PubMed=313486;
 RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
 RT "Differential amplification of antifreeze protein genes in the Pleuronectinae."
 RL J. Mol. Evol. 27:29-35(1988).
 RN [3]
 RP SEQUENCE FROM N.A. (PROTEIN B).
 RX MEDLINE=84264559; PubMed=6086629;
 RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
 RT "Antifreeze protein genes of the winter flounder."
 RL J. Biol. Chem. 259:9241-9247(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem repeats."
 RL Gene 112:163-170(1992).
 RN [5]
 RP 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148833; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding mechanism."
 RL J. Mol. Biol. 223:509-517(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RX MEDLINE=95281060; PubMed=7760940;
 RA Sichert F., Yang D.S.C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from winter flounder."
 RL Nature 375:427-431(1995).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L00138; AAB59964.1; -;
DR EMBL: L29178; AAB59964.1; JOINED.
DR EMBL: M62414; AAA49469.1; -;
DR EMBL: X07506; CAA30389.1; -;
DR EMBL: M62416; AAA49471.1; -;
DR EMBL: M62417; AAA49472.1; -;
DR PIR: A05161; A05161.
DR PIR: J50704; PDPLAN.
DR PIR: J50706; J50706.
DR PIR: S02326; S02326.
DR PDB: 1ATF; 15-OCT-94.
DR PDB: 1WFA; 03-JUN-95.
DR PDB: 1WFB; 03-JUN-95.
DR InterPro: IPR000104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
DR Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
FT SIGNAL 1 21
FT PROPEP 22 44
FT CHAIN 45 82 REMOVED BY A DIPEPTIDYLPEPTIDASE
FT VARIANT 36 36 (PROBABLY).
FT VARIANT 70 70 ANTIFREEZE PROTEIN A/B.
FT CONFLICT 24 24 A -> V.
FT HELIX 46 80 A -> D (IN PROTEIN B).
FT SEQUENCE 82 AA; 7711 MW; C2AB7B74C0D46CCL CRC64;
S -> R (IN REF. 2).
Query Match 3.2%; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred.No.15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126
DB 35 AAAAPAA 41

Search completed: January 15, 2004, 15:29:01
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:26:21 ; Search time 35 Seconds
(without alignments)
1614.672 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MGLLMACIVCAFRKRRF.....TAKPAPEHPSPSLQANQ 219

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size: 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	4.1	73	5 Q8MSH9	Q8MSH9 drosophila
2	9	4.1	75	5 Q9VBL6	Q9VBL6 drosophila
3	9	4.1	344	11 Q9DSJ3	Q9DSJ3 mus musculus
4	9	4.1	784	16 Q8PH14	Q8PH14 xanthomonas
5	9	4.1	841	5 Q9NEA5	Q9NEA5 leishmania
6	9	4.1	2969	16 Q8NS46	Q8NS46 corynebacte
7	8	3.7	125	10 Q8H4F8	Q8H4F8 oryza sativ
8	8	3.7	139	16 Q8Y2D2	Q8Y2D2 raietonia b
9	8	3.7	141	16 Q8Y2R3	Q8Y2R3 raietonia b
10	8	3.7	145	10 Q9EFC0	Q9EFC0 arabidopsis
11	8	3.7	145	10 Q96516	Q96516 arabidopsis
12	8	3.7	193	10 Q9C7F7	Q9C7F7 arabidopsis
13	8	3.7	193	10 Q8LE10	Q8LE10 arabidopsis
14	8	3.7	209	15 Q9DOU8	Q9DOU8 human immun
15	8	3.7	228	16 Q9AKA2	Q9AKA2 caulobacter
16	8	3.7	231	2 Q93ND2	Q93ND2 myxococcus

17	8	3.7	233	2 Q8GGE5	Q8GGE5 streptomyc
18	8	3.7	242	4 Q8NNG8	Q8NNG8 homo sapien
19	8	3.7	259	2 Q91UO4	Q91UO4 rhizobium m
20	8	3.7	262	10 Q8W2Q2	Q8W2Q2 oryza sativ
21	8	3.7	266	16 Q8P710	Q8P710 xanthomonas
22	8	3.7	268	16 P95192	P95192 mycobacteri
23	8	3.7	272	10 Q8S2S9	Q8S2S9 thellungiel
24	8	3.7	276	16 Q92WH2	Q92WH2 rhizobium m
25	8	3.7	285	11 Q8BHC2	Q8BHC2 mus musculu
26	8	3.7	291	10 Q40478	Q40478 nicotiana t
27	8	3.7	291	10 Q8H382	Q8H382 oryza sativ
28	8	3.7	302	5 Q9N9A5	Q9N9A5 leishmania
29	8	3.7	325	2 Q922X8	Q922X8 frateuria b
30	8	3.7	336	16 Q9AB24	Q9AB24 caulobacter
31	8	3.7	352	11 Q9CX00	Q9CX00 mus musculu
32	8	3.7	364	16 Q8B885	Q8B885 xanthomonas
33	8	3.7	360	11 Q8B2V6	Q8B2V6 mus musculu
34	8	3.7	381	2 Q86996	Q86996 acinetobact
35	8	3.7	384	16 Q9AD02	Q9AD02 streptomyc
36	8	3.7	401	8 Q8ME37	Q8ME37 primula jes
37	8	3.7	426	10 Q94108	Q94108 zea mays lm
38	8	3.7	434	10 Q8H3B5	Q8H3B5 oryza sativ
39	8	3.7	443	8 Q8ME30	Q8ME30 primula poi
40	8	3.7	448	8 Q31330	Q31330 streptitza
41	8	3.7	468	8 Q9GEC0	Q9GEC0 primula cor
42	8	3.7	468	8 Q8ME93	Q8ME93 primula bul
43	8	3.7	493	2 Q48431	Q48431 klebsiella
44	8	3.7	493	5 Q9KAP4	Q9KAP4 drosophila
45	8	3.7	493	5 Q8SZK1	Q8SZK1 drosophila

ALIGNMENTS

RESULT 1

Q8MSH9 ID Q8MSH9 PRELIMINARY; PRT; 73 AA.
AC Q8MSH9;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE GH19893P (Fragment).
GN BCDNA:GH19893.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dreesnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guartin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.B., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY118602; AAM50662.1; -
DR FlyBase; FBgn006325; BCDNA:GH19893.
FT NON TER 1
SQ SEQUENCE 73 AA; 6847 MW; BC8B7C976694F2D CRC64;

Query Match 4.1%; Score 9; DB 5; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAAP 128
Db 43 AAAAPAAP 51

RESULT 2

09YBL6
ID 09YBL6 PRELIMINARY; PRT; 75 AA.
AC 09YBL6, Q24390;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Accessory gland-specific peptide 57da precursor (Male accessory gland
secretory protein 57da).
GN M57DA OR CG9074.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=OREGON-R;
RX MEDLINE=95227188; PubMed=711745;
RA Stimmerl E., Schaefer M., Schaefer U.,
RT "Structure and regulation of a gene cluster for male accessory gland
transcript in Drosophila melanogaster."
RL Insect Biochem. Mol. Biol. 25:127-137(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Flamkoc C., Baldwin D.,
RA Bailett R.M., Basso P.V., Berman B.P., Bhandari D., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Beasley E.M.,
RA Borikava D., Botchan M.R., Butler H., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.S., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Miletina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING
CC MATING AND MAY AFFECT EGG-LAYING AND BEHAVIOR AFTER MATING.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: LUMEN FLUID OF MALE ACCESSORY GLANDS, BECOMES
CC SEMINAL FLUID.
DR EMBL: Z33647; AA03925.1; -;
DR EMBL: A5003753; AAF56515.1; -;
DR FlyBase: FBgn0011668; Maf57da.
KW Signal; Behavior.

FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 75 ACCESSORY GLAND-SPECIFIC PEPTIDE 57DA.
FT DOMAIN 39 64 ALA/PRO-RICH DOMAIN.
FT CONFLICT 39 46 MISSING (IN REF. 1).
FT CONFLICT 64 75 MISSING (IN REF. 1).
SQ SEQUENCE 75 AA; 7106 MW; BCD3D70817C98E14 CRC64;
Query Match 4.1%; Score 9; DB 11; Length 75;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 120 AAAAPAPP 128
Db 45 AAAAPAPP 53
RESULT 3
ID 09D5J3 PRELIMINARY; PRT; 344 AA.
AC 09D5J3;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 4930432K03R1X protein.
GN 4930432K03R1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OK NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawachi T., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochita H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Nombaretti P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitteker C., Wilming L.,
RA Wymahaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
DR EMBL: AK015291; BAB29782.1; -;
DR MGI: 1921029; 4930432K03R1X.
DR InterPro: IPR002965; P rich exten.
DR PRINTS: PR01217; PRICHTENSN.
SQ SEQUENCE 344 AA; 36294 MW; D65EAD71C5802AC CRC64;
Query Match 4.1%; Score 9; DB 11; Length 344;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGLIMACI 9
Db 1 MGLIMACI 9
RESULT 4
ID 08PH14 PRELIMINARY; PRT; 784 AA.
AC 08PH14;
DT 01-OCT-2002 (TRENBLrel. 22, Created)

```

DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)
DE Hypothetical protein XAC3446.
GN XAC3446.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Camargo F., Cardozo J., Chambeiro F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Medeiros A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Stuhel J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011991; AAM38289.1; -.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00486; trans_reg_C.1.
DR Prodom; PD000329; Trans_reg_C.1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 784 AA; 84093 MW; F41377005382EC8 CRC64;

QY 152 APVAAPAA 160
Db 160 APVAAPAA 168

Query Match 4.1%; Score 9; DB 16; Length 784;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
Q9NEAS PRELIMINARY; PRT; 841 AA.
AC Q9NEAS;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Possible inner dynein arm.
GN L5515.03.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=9816435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome."
RL Genome Res. 8:135-145(1998).
EMBL; AL161414; CAB7677.1; -.

```

```

DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 841 AA; 93125 MW; B41B72F1B8821226 CRC64;

Query Match 4.1%; Score 9; DB 5; Length 841;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAAP 127
Db 40 SAAAPAAP 48

RESULT 6
Q8NS46 PRELIMINARY; PRT; 2969 AA.
AC Q8NS46;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE 3-oxoacyl-(acyl-carrier-protein) synthase (EC 2.3.1.85).
GN CGL0836.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005276; BAB98229.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR003965; Fatty acid synth.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR002539; Maoc dehydratase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00698; Acyl_transf.1.
DR Pfam; PF00109; ketoacyl-synt.1.
DR Pfam; PF02801; ketoacyl-synt.1.
DR Pfam; PF01575; Maoc dehydratase.1.
DR PRINTS; PR01483; PASTINASE.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
KW Transferase; Acyltransferase; Complete proteome.
SQ SEQUENCE 2969 AA; 315124 MW; 908BA5A60D90815 CRC64;

Query Match 4.1%; Score 9; DB 16; Length 2969;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAPAPAA 176
Db 1663 PAAPAPAA 1671

RESULT 7
Q8H4F8 PRELIMINARY; PRT; 125 AA.
AC Q8H4F8;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE OJ1221 H04.16 protein.
GN OJ1221 H04.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

```

OC Ehirarioidae; Oryzae; Oryza.
OX NCB1_Taxid=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.,
RT "Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 8, BAC
  clone:O1221.H04."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004183; BAC21432.1; -.
SQ SEQUENCE 125 AA; 13560 MW; 31F75D839B8A52 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAAP 128
DB 11 AAAPAAP 18

RESULT 8
Q8Y2D2 PRELIMINARY; PRT; 139 AA.
AC Q8Y2D2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical signal peptide protein RSC0404.
GN RSC0404 OR RS03372.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCB1_Taxid=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1100;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
  Arlet M., Billault A., Brottier P., Camus J.C., Catolico L.,
  Chandler M., Choisme N., Claudel-Renard C., Cumac S., Demange N.,
  Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
  Signier P., Thebaud P., Whalen M., Winkler P., Levy M.,
  Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AF646059; CAD13932.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA; 14153 MW; A3150B6FD629039 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAAP 127
DB 55 AAAPAAP 62

RESULT 9
Q8Y2R3 PRELIMINARY; PRT; 141 AA.
AC Q8Y2R3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein RSC0272.
GN RSC0272 OR RS03239.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCB1_Taxid=305;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=GM1100;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
  Arlet M., Billault A., Brottier P., Camus J.C., Catolico L.,
  Chandler M., Choisme N., Claudel-Renard C., Cumac S., Demange N.,
  Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
  Signier P., Thebaud P., Whalen M., Winkler P., Levy M.,
  Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AF646058; CAD13800.1; -.
DR InterPro; IPR001763; Rhodanese-like.
DR SMART; SM00450; RHOD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 15454 MW; 878FE1840A8E8B4E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156
DB 126 AAGAPVAA 133

RESULT 10
Q9PFC0 PRELIMINARY; PRT; 145 AA.
AC Q9PFC0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Histone H2B like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCB1_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
  Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
  features of the 1.6 Mb regions covered by twenty physically assigned
  BAC clones."
RL DNA Res. 4:215-230(1997).
CC CC -1 SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
  H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
  SIMILARITY).
CC CC -1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC CC -1 SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
DR EMBL; AB005243; BAB10609.1; -.
DR InterPro; IPR004822; Histone core.
DR InterPro; IPR000558; Histone_H2B.
DR Pfam; PF00125; histone; 1.
DR PRINTS; PR00621; HISTONEH2B.
DR ProDom; PD000497; Histone_H2B; 1.
DR SMART; SM00427; H2B; 1.
DR PROSITE; PS00357; HISTONE_H2B; 1.
KW Chromosomal protein; DNA-Binding; Nuclear protein; Nucleosome core.
SQ SEQUENCE 145 AA; 15732 MW; CC6421B559D42B30 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AABPAAA 137
DB 19 AABPAAA 26

```



```

RESULT 11
ID 096516 PRELIMINARY; PRT; 145 AA.
AC 096516;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Histone H2B like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ecotype columbia;
RA Phillips G.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1 SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
CC SIMILARITY).
CC -1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC EMBL; Y07745; CA69025.1; -.
DR InterPro; IPR004822; Histone_core.
DR InterPro; IPR000558; Histone_H2B.
DR Pfam; PF00125; histone; 1.
DR PRINTS; PR00621; HISTONEH2B.
DR PRODOM; PD000497; Histone_H2B; 1.
DR SMART; SM00427; H2B; 1.
DR PROSITE; PS00357; HISTONE_H2B; 1.
KM Chromosomal protein; DNA-binding; Nuclear protein; Nucleosome core.
SQ SEQUENCE 145 AA; 1573 MW; 6C6E8B18390F4666 CRC64;

Query Match 3.7%; Score 8; DB 10; Length 145;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAAPAAA 137
Db 19 AAAPAAA 26

RESULT 12
ID 09C7F7 PRELIMINARY; PRT; 193 AA.
AC 09C7F7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Lipid transfer protein, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA MEDLINE=21016719; PubMed=11130712;
RA Theologis A.; Ecker J.R.; Palm C.J.; Federapfel N.A.; Kaul S.;
RA White O.; Alonso J.; Altafaj H.; Araujo R.; Bowman C.L.; Brooks S.Y.;
RA Buehler E.; Chan A.; Chao Q.; Chen H.; Cheuk R.F.; Chin C.W.;
RA Chung M.K.; Conn L.; Conway A.B.; Creasy T.H.; Dewar K.;
RA Dunn P.; Etgu P.; Feldblum T.V.; Feng J.-D.; Fong B.; Fujii C.Y.;
RA Gali J.E.; Goldsmith A.D.; Haas B.; Hansen N.F.; Hughes B.; Hultzar L.;
RA Hunter J.L.; Jenkins J.; Johnson-Hopson C.; Khan S.; Khaykin E.;
RA Kim C.J.; Koo H.L.; Kremenetskaia I.; Kurtz D.B.; Kwan A.; Lam B.;
RA Langin-Hooper S.; Lee A.; Lee J.M.; Lenz C.A.; Li J.H.; Li Y.-P.;
RA Lin X.; Liu S.X.; Liu Z.A.; Luros J.S.; Malti R.; Marziani A.;
RA Miltscher J.; Miranda M.; Nguyen M.; Niernan W.C.; Osborne B.I.;

```

```

RA Pai G.; Peterson J.; Pham P.K.; Rizzo M.; Rooney T.; Rowley D.;
RA Sakano H.; Salzer S.L.; Schwartz J.R.; Shinn P.; Southwick A.M.;
RA Sun R.; Tallon L.J.; Tambunga G.; Toriumi M.J.; Town C.D.;
RA Utecherack T.; Van Aken S.; Vaysberg M.; Vysotskaya V.S.; Walker M.;
RA Wu D.; Yu G.; Fraser C.M.; Venter J.C.; Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A.; Karlin-Neumann G.; Nguyen M.; Lam B.; Miranda M.;
RA Palm C.J.; Bowser L.; Jones T.; Banb J.; Carninci P.; Chen H.;
RA Cheuk R.; Chung M.K.; Hayashizaki Y.; Ishida J.; Kamuya A.; Kawai J.;
RA Kim C.; Lin J.; Liu S.X.; Narusaka M.; Pham P.K.; Sakano H.;
RA Sakurai T.; Satou M.; Seki M.; Shinn P.; Yamada K.; Shinozaki K.;
RA Becker J.; Theologis A.; Davis R.W.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Tripp M.; Southwick A.; Karlin-Neumann G.; Nguyen M.; Miranda M.;
RA Palm C.J.; Bowser L.; Jones T.; Banb J.; Carninci P.; Chen H.;
RA Cheuk R.; Chung M.K.; Hayashizaki Y.; Ishida J.; Kamuya A.; Kawai J.;
RA Kim C.; Lin J.; Liu S.X.; Narusaka M.; Pham P.K.; Sakano H.;
RA Sakurai T.; Satou M.; Seki M.; Shinn P.; Yamada K.; Shinozaki K.;
RA Becker J.; Theologis A.; Davis R.W.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC069471; AGS1485.1; -.
DR EMBL; AY092956; AAM12955.1; -.
DR EMBL; AY128712; AAM91112.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tryd_alpha_amy1; 1.
SQ SEQUENCE 193 AA; 19759 MW; D54B38B12FF6610 CRC64;

Query Match 3.7%; Score 8; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAPAAP 127
Db 20 AAAPAAP 27

RESULT 13
ID 08LE10 PRELIMINARY; PRT; 193 AA.
AC 08LE10;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Lipid transfer protein, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J.; Volfovsky N.; Town C.D.; Troukhan M.; Alexandrov N.;
RA Feldmann K.A.; Flavell R.B.; White O.; Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation."
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V.; Troukhan M.; Alexandrov N.; Lu Y.-P.; Flavell R.;
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY085407; AAM62634.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tryd_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 193 AA; 19747 MW; C091074205D76610 CRC64;

```

Query Match 3.7%; Score 8; DB 10; Length 193;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 DB 20 AAAAPAP 27

RESULT 14

Q9D0U8 PRELIMINARY; PRT; 209 AA.
 AC Q9D0U8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Nef protein (Negative factor) (27 kDa protein).
 GN Nef.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviridae; Lentivirus.
 OC NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=01u09;
 RX MEDLINE=21002575; PubMed=11118071;
 RA Geffin R., Wolf D., Miller R., Hill M.D., Stellwag B., Freitag M.,
 RA Saks G., Scott G.B., Baur A.S.;
 RT "Functional and structural defects in HIV-1 nef genes derived from
 RT pediatric long-term survivors."
 RL AIDS Res. Hum. Retroviruses 16:1855-1868(2000).
 CC -1- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
 CC ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4 (T4) ANTIGEN (BY
 CC SIMILARITY).
 DR EMBL; AF252901; AAC34573.1; -.
 DR HSP; P03406; IEFN.
 DR InterPro; IPR001558; HIV_Nef.
 DR Pfam; PF00469; F-protein; 1.
 DR ProDom; PD000031; HIV_Nef; 1.
 KM AIDS; GTP-binding; Lipoprotein; Myristate.
 SQ SEQUENCE 209 AA; 23656 MW; 6B56B8AD98C79519 CRC64;

Query Match 3.7%; Score 8; DB 15; Length 209;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 ABPAPAP 138
 DB 23 ABPAPAP 30

RESULT 15

Q9A4A2 PRELIMINARY; PRT; 228 AA.
 AC Q9A4A2;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cytochrome c, membrane-bound.
 GN CC2935.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
 OC Caulobacteraceae; Caulobacter.
 OC NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Knouri H., Shetty J., Berry K.,

RA Utterback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AB005958; AK24897.1; -.
 DR HSP; P01458; I154.

DR TIGR; CC2935; -.
 DR InterPro; IPR000104; Antifreeze_1.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002327; Cyt_C1AB.
 DR Pfam; PF000034; cytochrome_c; 1.
 DR PRINTS; PR00308; ANTI-FREEZE1.
 DR PRINTS; PR00604; CYTCHROME_C1AB.
 DR ProDom; PD000375; Cyt_C1AB; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KM Complete proteome.
 SQ SEQUENCE 228 AA; 23023 MW; 5AB05C7DD5D05F82 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 228;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 DB 219 AAAAPAP 226

Search completed: January 15, 2004, 15:29:49
 Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:26:56 ; Search time 21 Seconds

(without alignments)
1002.902 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLMAGIVCAVAFARRRRF.....TAKPADEPHPSLSQANQ 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: 1: PIR.76:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.1	581	2	probable glutamyl-nitrate reductase
2	9	4.1	864	1	JC1422
3	8	3.7	110	1	R6BYP3
4	8	3.7	129	2	H86500
5	8	3.7	129	2	C72122
6	8	3.7	193	2	H86404
7	8	3.7	228	2	H87612
8	8	3.7	268	2	G70645
9	8	3.7	276	2	G95887
10	8	3.7	291	2	T02434
11	8	3.7	325	2	T48873
12	8	3.7	336	2	A67300
13	8	3.7	353	2	S41958
14	8	3.7	381	2	T46827
15	8	3.7	396	1	JH0633
16	8	3.7	495	2	T52066
17	8	3.7	506	2	A40679
18	8	3.7	523	2	B40679
19	8	3.7	550	2	C75557
20	8	3.7	574	2	S23530
21	8	3.7	611	2	C96030
22	8	3.7	846	2	S52418
23	8	3.7	1476	2	A45773
24	8	3.7	4307	2	T20721
25	8	3.2	78	2	S33173
26	8	3.2	38	2	B83446
27	7	3.2	82	1	FDPLAW
28	7	3.2	82	1	S02326
29	7	3.2	82	2	A05161

30	7	3.2	82	2	T51125	antifreeze protein
31	7	3.2	100	2	F58552	hypothetical prote
32	7	3.2	101	2	H69091	ribosomal protein
33	7	3.2	102	2	D64363	ribosomal protein
34	7	3.2	106	1	R6DOP2	acidic ribosomal p
35	7	3.2	107	1	R6KMLC	acidic ribosomal p
36	7	3.2	109	1	R6JUMP1	acidic ribosomal p
37	7	3.2	110	2	T37490	ribosomal protein
38	7	3.2	111	2	E72524	probable ribosomal
39	7	3.2	111	2	B95857	hypothetical prote
40	7	3.2	112	1	R5FP28	acidic ribosomal p
41	7	3.2	112	2	S54179	acidic ribosomal p
42	7	3.2	112	2	E86141	protein T25K16.9
43	7	3.2	114	1	RSRT12	acidic ribosomal p
44	7	3.2	114	1	RSBS2H	ribosomal protein
45	7	3.2	114	2	F84266	50S ribosomal prot

ALIGNMENTS

RESULT 1

probable glutamyl-tRNA reductase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #Sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T36267

R/Murphy, L.; Harries, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, June 1999

A/Reference number: Z21576

A/Accession: T36267

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-581 <MUR>

A/Experimental source: strain A3(2)

C/Genetics:

A/Genes: SCQDB:SCB68.17C

Query Match

Best Local Similarity 100.0%; Pred. No. 2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 154 VAAEPAAEA 162

36 VAAEPAAEA 44

RESULT 2

JC1422

nitrate reductase (NADH) (EC 1.7.1.1) - Volvox carter1

C/Species: Volvox carter1

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002

C/Accession: JC1422; S22192

R/Gruber, H.; Goeltink, S.D.; Kirk, D.L.; Schmitt, R.

Gene 120, 75-83, 1992

A/Title: The nitrate reductase-encoding gene of Volvox carter1: Map location, sequence

A/Reference number: JC1422; MID:93013022; PMID:1398126

A/Accession: JC1422

A/Molecule type: DNA

A/Residues: 1-864 <GR2>

A/Cross-references: EMBL:X64136; NID:g21993; PTDN:CAA45497.1; PTD:g21994

A/Note: submitted to the EMBL Data Library, January 1992

C/Genetics:

A/Map position: linkage group IX

A/Insertions: 183/3; 234/2; 295/1; 339/1; 425/2; 521/3; 593/3; 677/3; 797/2

C/Function: catalyzes the reduction of nitrate to nitrite

A/Description: nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 red

C/Keywords: dimer; electron transfer; FAD; flavoprotein; heme; iron; metalloprotein; mol

F;36-425/Domain: molybdopterin-binding domain homology <PCO>

F;497-571/Domain: cytochrome b5 core homology <CB5>

F/613-864/Domain: cytochrome-b5 reductase homology <CBR>
F/139/Binding site: molibdopterin (Cys) (covalent) #status predicted
F/532/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
DB 478 AAAAPAAP 486

RESULT 3

606 acidic ribosomal protein p1-alpha - fission yeast (*Schizosaccharomyces pombe*)
N/Alternate names: acidic ribosomal protein 3; ribosomal protein Y122II
C/Species: *Schizosaccharomyces pombe*
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 10-Dec-1999
R/Accession: C34715; T40352
R/Beltrame, M.; Bianchi, M.E.
Mol. Cell. Biol. 10, 2341-2348, 1990
A/Title: A gene family for acidic ribosomal proteins in *Schizosaccharomyces pombe*: two
A/Reference number: A34715; MUID:90220620; PMID:2325655

A/Accession: C34715
A/Molecule type: DNA
A/Residues: 1-110
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
R/Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z21922

A/Accession: T40352
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-110 <NOO>
A/Cross-references: EMBL:AL022070; PIDD:CAA17793.1; GSPDB:GN00067; SPDB:SPBC39.13C
A/Experimental source: strain 972h-; cosmid C3B9

C/Genetics:
A/Gene: rpa3
A/Map position: 2
A/Intons: 23/3; 103/3
C/Supfamily: rat acidic ribosomal protein p1
C/Keywords: phosphoprotein; protein biosynthesis; ribosome

Query Match 3.7%; Score 8; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAAAP 126
DB 64 SAAAAAP 71

RESULT 4

L7/L12 ribosomal protein [imported] - *Chlamydomonas reinhardtii* (strain J138)
C/Species: *Chlamydomonas reinhardtii*, *Chlamydomonas reinhardtii*
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
R/Accession: H86500
R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A/Title: Comparison of whole genome sequences of *Chlamydomonas reinhardtii* J138.
A/Reference number: A86491; MUID:20330349; PMID:10871362

A/Accession: H86500
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-129 <STO>
A/Cross-references: GB:BA000006; NID:g8978453; PIDD:BA98290.1; GSPDB:GN00142
A/Experimental source: strain J138
C/Genetics:
A/Gene: r17

C/Supfamily: *Escherichia coli* ribosomal protein L12

Query Match 3.7%; Score 8; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 BAPVAP 178
DB 50 BAPVAP 57

RESULT 5

ribosomal protein L7/L12 CP0695 [imported] - *Chlamydomonas reinhardtii* (strain CWL029)
C/Species: *Chlamydomonas reinhardtii*, *Chlamydomonas reinhardtii*
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C/Accession: C72122; G81548
R/Kalman, S.; Mitchell, W.; Marathe, R.; Lamell, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A/Title: Comparative genomes of *Chlamydomonas reinhardtii* and *C. trichomatis*.
A/Reference number: A72000; MUID:99206606; PMID:10192388

A/Accession: C72122
A/Molecule type: DNA
A/Residues: 1-129 <ARN>
A/Cross-references: GB:AE001593; GB:AE001363; NID:g4376334; PIDD:AA018233.1; PID:g43763

A/Experimental source: strain CWL029
R/Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of *Chlamydomonas trichomatis* Mopn and *Chlamydomonas reinhardtii* AR39.
A/Reference number: A81500; MUID:20150255; PMID:10684935

A/Accession: G81548
A/Molecule type: DNA
A/Residues: 1-129 <REA>
A/Cross-references: GB:AE002228; GB:AE002161; NID:g7189606; PIDD:AA038503.1; PID:g71896
A/Experimental source: strain AR39, HL cells
C/Genetics:
A/Gene: r17; CP0695

C/Supfamily: *Escherichia coli* ribosomal protein L12

Query Match 3.7%; Score 8; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 BAPVAP 178
DB 50 BAPVAP 57

RESULT 6

H86404
probable lipid transfer protein [imported] - *Arabidopsis thaliana*
C/Species: *Arabidopsis thaliana* (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C/Accession: H86404

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Hulzer, L.
Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marshall
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: H86404
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-193 <STO>
A/Cross-references: GB:AE005172; NID:g11024871; PIDD:AA026955.1; GSPDB:GN00141
C/Genetics:
A/Map position: 1

Query Match 3.7%; Score 8; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
Db 20 AAAAPAP 27

RESULT 7

E87612

Cytochrome c, membrane-bound [imported] - *Caulobacter crescentus*

C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 18-Nov-2002

C:Accession: E87612

R:Name: W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Lamb, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Klotzel, J.; Krumholz, L.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87612

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-228 <STO>

A:Cross-references: GB:A805673; NID:913424561; PIDN:AAK24897.1; GSPDB:GN00148

A:Gene: CC2935

C:Superfamily: membrane-bound cytochrome c/cytochrome c homology

C:Keywords: chromoprotein; heme; iron; metalloprotein

F:81/84/Binding site: heme (Cys) (covalent) #status predicted

F:85/Binding site: heme iron (His) (axial ligand) #status predicted

F:150/Binding site: heme iron (Met) (axial ligand) #status predicted

Query Match 3.7%; Score 8; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
Db 219 AAAAPAP 226

RESULT 8

G70645
Hypothetical protein RV3134C - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70645

R:Name: S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70645

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-268 <COL>

A:Cross-references: GB:283667; GB:AL123456; NID:93261655; PIDN:CAB06280.1; PID:4290951;

A:Experimental source: strain H37RV

C:Gene: RV3134C

Query Match 3.7%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 PSPAEPAT 200
Db 138 PSPAEPAT 145

RESULT 9
G95887
Probable ABC transporter permease protein SMD20381 [imported] - *Sinorhizobium meliloti*

C:Species: *Sinorhizobium meliloti*

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 01-Mar-2002

C:Accession: G95887

R:Name: T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chait, P.; Vorholter, F.J.; Herra,

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-kb *psymb* megaplasmid from the N2-fixing end

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95887

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <RUC>

A:Cross-references: GB:AL591985; PIDN:CAC48767.1; PID:915140240; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid *psymb*

R:Galibert, F.; Pinan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubier

pelle, D.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure

heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Content: annotation

C:Gene: SMD20381

A:Gene: SMD20381

A:Superfamily: spermidine/putrescine transport system permease protein port

Query Match 3.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAAPAAA 137
Db 268 AAAPAAA 275

RESULT 10

T02434

DNA binding protein ERBP-4 - common tobacco

C:Species: *Nicotiana tabacum* (common tobacco)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 31-Mar-2001

C:Accession: T02434

R:Name: Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene response

A:Reference number: Z14671; MUID:95276459; PMID:7756828

A:Accession: T02434

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-291 <OHM>

A:Cross-references: EMBL:D38125; NID:9790361; PIDN:BA07323.1; PID:91208497

A:Experimental source: strain B74; tissue-type leaf

C:Superfamily: Arabidopsis thaliana hypothetical protein T19P19.170

Query Match 3.7%; Score 8; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 AAAPAPT 140
Db 247 AAAPAPT 254

RESULT 11

T48873

electron transfer protein [imported] - *Fraxaura* sp. (strain ANA-18)

C:Species: *Fraxaura* sp.

A:Variety: strain ANA-18

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000

C:Accession: T48873

R;Murakami, S.; Takashima, A.; Takemoto, J.; Takenaka, S.; Shinke, R.; Aoki, K.
Gene 226, 189-198, 1999

A/Title: Cloning and sequence analysis of two catechol-degrading gene clusters from the
A/Reference number: Z24833

A/Accession: T46873

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-325 <MUR>

A/Cross-references: EMBL:AB009373; PIDN:BA075213.1

A/Experimental source: strain ANA-18

C/Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 325;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAAPAAA 137
DB 223 AAAPAAA 230

RESULT 12

TPR domain protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C/Accession: A87300

R;Nierman, W.C.; Feildlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: A87300

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-336 <STO>

A/Cross-references: GB:AE005673; NID:G13421571; PIDN:AAK2397.1; GSPDB:GND0148

C/Genetics:

A/Gene: CC0410

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 336;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAAP 127
DB 112 AAAPAAP 119

RESULT 13

pupal cuticle protein precursor - greater wax moth

C/Species: Galleria mellonella (greater wax moth)

C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 19-May-2000

C/Accession: S41958

R;Kolberg, U.; Obermeyer, B.; Hirsch, H.; Kelber, G.; Wolbert, P.

submitted to the EMBL Data Library, February 1994

A/Description: Expression cloning and characterization of a pupal cuticle protein cDNA C

A/Reference number: S41958

A/Accession: S41958

A/Molecule type: mRNA

A/Residues: 1-353 <ROL>

A/Cross-references: EMBL:X77514; NID:G453403; PIDN:CA54650.1; PID:G453404

F/1-15/Domain: signal sequence #status predicted <SIG>

F/16-353/Product: pupal cuticle protein #status predicted <MAT>

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 353;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAPAAA 126
|||||

DB 236 SAAPAAA 243

RESULT 14

T46827

phenoxycarboxylate dioxygenase [imported] - Acinetobacter lwoffii

C/Species: Acinetobacter lwoffii

C/Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000

C/Accession: T46827

R;Kim, S.I.; Leem, S.H.; Choi, J.S.; Chung, Y.H.; Kim, S.; Park, Y.M.; Ha, K.S.

submitted to the EMBL Data Library, November 1996

A/Reference number: Z24100

A/Accession: T46827

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-381 <KIM>

A/Cross-references: EMBL:U77659; PIDN:AA031770.1

A/Experimental source: strain K24

C/Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 381;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAAPAAA 137
DB 279 AAAPAAA 286

RESULT 15

JH0633

cellular tumor antigen p53 - golden hamster

N/Alternate names: tumor-suppressor protein p53

C/Species: Mesocricetus auratus (golden hamster)

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C/Accession: JH0633

R;Legros, Y.; McInyre, P.; Soussi, T.

Gene 112, 247-250, 1992

A/Title: The cDNA cloning and immunological characterization of hamster p53.

A/Reference number: JH0633; MUID:92210007; PMID:1555773

A/Accession: JH0633

A/Molecule type: mRNA

A/Residues: 1-356 <LEG>

A/Cross-references: GB:M75144; NID:G191414; PIDN:AAA37085.1; PID:G191415

A/Experimental source: kidney, strain MP1

C/Genetics:

A/Gene: p53

C/Superfamily: cellular tumor antigen p53

C/Keywords: apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosf

F/179,182,241,245/Binding site: zinc (Cys, His, Cys, Cys) #status predicted

F/395/Binding site: phosphoryl-RNA (Ser) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 8; DB 1; Length 396;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAAP 127
DB 66 AAAPAAP 73

Search completed: January 15, 2004, 15:30:23
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:07:29 ; Search time 41 Seconds

(without alignments)
847.832 Million cell updates/sec

Title: US-09-923-236-2.

Perfect score: 1178
Sequence: 1 MKLLMACIVCAFARRRRF.....TAKPADEPHPSLEGANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1178	100.0	219	22	AAB60109 Human transport pr
2	1178	100.0	219	23	AAE27862 Human zslg63 prote
3	1178	100.0	219	23	ABG31608 Human secreted gal
4	1178	100.0	219	23	AAU74536 Human zslg63 polyp
5	1178	100.0	219	24	ABU08515 Human zslg63 prote
6	1172	99.5	219	20	AAW30653 Human secreted pro
7	1113.5	94.5	221	20	AAI9472 Amino acid sequenc
8	113.5	15.6	325	22	ABG21919 Novel human diagn
9	180	15.3	207	14	AAK40797 Sequence of a 10-a

10	179	15.2	207	17	AAW06915
11	179	15.2	207	20	AAI32850
12	179	15.2	207	20	AAI23305
13	175.5	14.9	718	22	AAW42089
14	175.5	14.9	903	23	ABP41277
15	175.5	14.9	3177	22	AAW40303
16	174	14.8	566	22	ABE61040
17	173	14.7	538	22	AAE82806
18	173	14.7	1427	23	ABP63024
19	171.5	14.6	693	23	ABP69529
20	170.5	14.5	149	22	AAU00451
21	170	14.4	267	19	AAW81726
22	170	14.4	267	19	AAW64359
23	170	14.4	267	20	AAI39156
24	170	14.4	267	20	AAI39013
25	170	14.4	428	23	AAW50963
26	170	14.4	428	23	ABP70532
27	170	14.4	842	22	ABE66631
28	170	14.4	842	22	ABE71319
29	170	14.4	1028	22	ABU70689
30	170	14.4	3151	24	ABR39571
31	170	14.4	3173	23	ABR90743
32	170	14.4	3173	24	ABU54450
33	170	14.4	3176	24	ABR39570
34	169.5	14.4	783	19	AAW37151
35	169.5	14.4	787	19	AAW37152
36	169.5	14.4	802	19	AAW37153
37	169.5	14.4	802	22	AAU09139
38	169	14.3	299	22	ABE61276
39	168.5	14.3	550	22	ABE82807
40	168.5	14.3	706	22	ABE66633
41	165.5	14.0	763	18	AAW31852
42	164	13.9	572	18	AAW31855
43	163.5	13.9	214	17	AAE86913
44	162.5	13.8	471	22	ABE61785
45	162	13.8	1064	22	ABE66652

ALIGNMENTS

RESULT 1					
AAB60109 standard; Protein; 219 AA.					
XX	XX	XX	XX	XX	XX
AC	AAB60109;				
XX	28-MAR-2001 (first entry)				
DT	Human transport protein TPPT-29.				
XX					
DE	Human; transport protein; TPPT; transport disorder; metabolic disorder;				
XX					
KW	neurological disorder; cardiovascular disorder; reproductive disorder;				
KW	immune disorder; cancer.				
XX					
OS	Homo sapiens.				
XX					
FN	W0200078953-A2.				
XX					
PD	28-DEC-2000.				
XX					
PF	16-JUN-2000; 2000MO-US16668.				
XX					
PR	17-JUN-1999; 99US-0139923.				
PR	10-AUG-1999; 99US-0148177.				
PR	18-AUG-1999; 99US-0149357.				
PR	28-OCT-1999; 99US-0162287.				
XX					
PA	(INCY-) INCYTE GENOMICS INC.				
XX					
PI	Lai P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;				
XX	Baughn MR, Azimzai Y, Lu DM, Au-Young J, Patterson C;				
XX					

T. cruzi Tcd anti
Tcd protein sequen
Tcd antigen of TTP
Human polypeptide
Human polypeptide
Human ovarian anti
Human polypeptide
Drosophila melanog
Human low density
Human polypeptide
Human polypeptide
Protein encoded by
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Maize methyl Cpg b
Histone deacetylase
Drosophila melanog
Drosophila melanog
Human adipocyte Se
Human alpha3 colla
Human Tumour Endoc
Human tumour endot
Human alpha3 colla
Mouse neural Mena+
Mouse neural Mena+
Mouse neural Mena+
Mammalian enabed
Drosophila melanog
Rabbit low density
Drosophila melanog
Mycobacterium tube
Mycobacterium tube
Cotton fiber-speci
Drosophila melanog
Drosophila melanog

DR MPI; 2001-041424/05.
 DR N-PSDB; AAF27729.
 XX Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 XX
 PS Claim 2; Page 130; 165pp; English.
 XX
 CC The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated rpts). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 1178; DB 22; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2, 5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLLMACIVCAFAKRRPPPIGDDNDGPHLSMIPGIRLPPPLYRPNVTP 60
 DB 1 MLLLMACIVCAFAKRRPPPIGDDNDGPHLSMIPGIRLPPPLYRPNVTP 60
 QY 61 SYPGNTYTDGLPSYPMILTSRPGFYVYHIRGFLATOLNVPPLPRGFPVPPSRFPA 120
 DB 61 STPGNTYTDGLPSYPMILTSRPGFYVYHIRGFLATOLNVPPLPRGFPVPPSRFPA 120
 QY 121 AAAPAAPPIAEPAAAPLTATPVAAPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180
 DB 121 AAAPAAPPIAEPAAAPLTATPVAAPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180
 QY 181 EAPVGVPAAPAEPSPAEPATAKPAAPBPSPSLDAQ 219
 DB 181 EAPVGVPAAPAEPSPAEPATAKPAAPBPSPSLDAQ 219
 QY 181 EAPVGVPAAPAEPSPAEPATAKPAAPBPSPSLDAQ 219
 DB 181 EAPVGVPAAPAEPSPAEPATAKPAAPBPSPSLDAQ 219
 RESULT 2
 ID AAE27862 standard; Protein; 219 AA.
 AC AAE27862;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 DE Human zsig63 protein.
 XX
 XX Human; secreted salivary protein; zsig63 protein; host defense protein;
 KM immune modulating factor; antipathogenic; cell-cell signalling molecule;
 KM growth factor; cytokine; growth factor hormone activity; dental carrier;
 KM infection; tooth decay; periodontal disease; gastrointestinal disease;
 KM through; urinary tract infection; vaginal infection; diabetes; obesity;
 KM anti-inflammatory; chronic tissue damage; lung dysfunction; testostis;
 KM gene therapy; salivary gland dysfunction; prostate gland dysfunction;
 KM forensic DNA profiling; chondrosarcoma; atherosclerosis; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 1..15
 FT Peptide /label= "Signal_peptide"
 FT 16..219
 FT Protein /note= "Mature human zsig63 protein"
 FT 14..21
 FT Region /note= "Hydrophilic region"
 FT 16..37
 FT Domain /note= "Domain 1"
 FT 17..33
 FT Region /note= "Antigenic epitope"
 FT 24..30
 FT Region /note= "Hydrophilic region"
 FT 38..126
 FT Domain

FT Region /note= "Domain 2"
 FT 66..73
 FT /note= "Antigenic epitope"
 FT 103..108
 FT /note= "Antigenic epitope"
 FT 124..133
 FT /note= "Repeat 1"
 FT 127..219
 FT /note= "Domain 3"
 FT 134..138
 FT /note= "Repeat 2"
 FT 139..143
 FT /note= "Repeat 3"
 FT 144..148
 FT /note= "Repeat 4"
 FT 149..153
 FT /note= "Repeat 5"
 FT 154..158
 FT /note= "Repeat 6"
 FT 159..163
 FT /note= "Repeat 7"
 FT 164..168
 FT /note= "Repeat 8"
 FT 169..173
 FT /note= "Repeat 9"
 FT 174..178
 FT /note= "Repeat 10"
 FT 179..183
 FT /note= "Repeat 11"
 FT 184..188
 FT /note= "Repeat 12"
 FT 187..192
 FT /note= "Hydrophilic region"
 FT 189..193
 FT /note= "Repeat 13"
 FT 190..197
 FT /note= "Antigenic epitope"
 FT 194..198
 FT /note= "Repeat 14"
 FT 199..203
 FT /note= "Repeat 15"
 FT 202..215
 FT /note= "Antigenic epitope"
 FT 204..208
 FT /note= "Repeat 16"
 XX
 XX
 XX US2002090677-A1.
 XX
 XX 11-JUL-2002.
 XX
 XX 03-AUG-2001; 2001US-0923236.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX 17-MAR-2000; 2000US-0527345.
 XX
 XX (ADLER/) ADLER D A.
 XX (SHEP/) SHEPPARD P O.
 XX
 XX
 XX Adler DA, Sheppard PO;
 XX
 XX MPI; 2002-642378/69.
 XX N-PSDB; AAD45050.
 XX
 PT Novel secreted salivary polypeptide, zsig63, useful as antimicrobial
 PT agent for treating microbial infection, dental carrier, periodontal
 PT disease, through gastrointestinal disease, and for aiding digestion -
 XX
 PS Claim 10; Page 28; 33pp; English.
 XX
 XX The invention relates to human secreted salivary polypeptide designated
 CC as zsig63 and nucleic acid molecules encoding such polypeptides. zsig63
 CC can be used in detecting agonists and antagonists of its activity, and
 CC is also useful as a host defense polypeptide, immune modulating factor,

CC antipathogenic polypeptide, cell-cell signalling molecule, growth factor,
 CC cytokine, or as secreted extracellular matrix associated proteins with
 CC growth factor hormone activity. It is useful for treating conditions
 CC associated with pathological microbes, including bacterial, fungal and
 CC viral infections, for treating dental caries (tooth decay), periodontal
 CC disease, thrush and gastrointestinal disease, for treating urinary tract
 CC infection, vaginal infection and for preventing infection in skin and
 CC other epithelial wounds. zsig63 is useful for establishing normal
 CC microflora and protect against pathogenic colonisation and invasion, for
 CC treating chronic tissue damage e.g. damage in extremities associated
 CC with diabetes and useful as anti-inflammatory agents. It is useful as a
 CC marker of lung dysfunction, salivary gland dysfunction, or dysfunction of
 CC prostate gland. It is also therapeutically useful for aiding digestion.
 CC Polynucleotides of the invention are used in gene therapy for increasing
 CC or inhibiting zsig63 activity, for detecting abnormalities on human
 CC chromosome 4 associated with disease or other human traits and as
 CC diagnostics in forensic DNA profiling. Sequences of the invention are
 CC useful for stimulating proliferation or differentiation of cardiac
 CC myocytes, for proliferation or differentiation of adipocytes and for
 CC inhibiting chondrosarcomas, atherosclerosis, restenosis and obesity.
 CC The present sequence is human zsig63 protein. zsig63 gene is located
 CC on chromosome 4.

CC Sequence 219 AA:

Query Match 100.0%; Score 1176; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACTIVCAAPKRRPPIGDDNDGHPILPSINIVYGRNLPPILYRPNVNP 60
 Db 1 MKLLIMACTIVCAAPKRRPPIGDDNDGHPILPSINIVYGRNLPPILYRPNVNP 60
 QY 61 SYPGNTYDTGLPSYPIWLTSPGFYVYHNGFPPLATOLNPPLPPRPGFPVPPSRFPSA 120
 Db 61 SYPGNTYDTGLPSYPIWLTSPGFYVYHNGFPPLATOLNPPLPPRPGFPVPPSRFPSA 120
 QY 121 AAAPAPPIAEPAAAPLVTATPVAAEPAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
 Db 121 AAAPAPPIAEPAAAPLVTATPVAAEPAGAPVAAEPAAAPVGAEPAAAPVAAEPAA 180
 QY 181 EAPVGEPAEPPAPAPATKAPAPRPPSPSLAQANQ 219
 Db 181 EAPVGEPAEPPAPAPATKAPAPRPPSPSLAQANQ 219

RESULT 3

ABG31608
 ID ABG31608 standard; Protein; 219 AA.

XX AC ABG31608;

DT 15-NOV-2002 (first entry)

XX DB Human secreted salivary protein zsig63.

XX KW Human; secreted salivary protein; zsig63; immunogen; zsig63-cytokine;
 KW antibody-cytokine; in vivo killing; pathological microbe; bacteria;
 KW fungal; viral; infection; salivary gland; anti-microbial; dental caries;
 KW tooth decay; periodontal disease; thrush; gastrointestinal disease;
 KW urinary tract infection; vaginal infection; skin infection; microflora;
 KW epithelial wound; pathogenic colonisation; invasion; pro-inflammatory;
 KW chronic tissue damage; vascular system; diabetes; anti-inflammatory;
 KW incompetent immune system; AIDS; acquired immunodeficiency syndrome;
 KW chemotherapy; radiation treatment; lung infection; cystic fibrosis;
 KW digestion; chromosome 4.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Region 1..15

FT /note= "Fusion protein peptide, specifically claimed
 FT in claim 18"

FT Region
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 9..204
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 14..19
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 16..21
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 16..37
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 16..219
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 17..33
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 24..29
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 24..33
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 25..30
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 38..126
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 66..73
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 103..108
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 109..215
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 124..133
 FT /label= Repeat_1
 FT 127..219
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 134..138
 FT /label= Repeat_2
 FT 139..143
 FT /label= Repeat_3
 FT 144..148
 FT /label= Repeat_4
 FT 149..153
 FT /label= Repeat_5
 FT 154..158
 FT /label= Repeat_6
 FT 159..163
 FT /label= Repeat_7
 FT 164..168
 FT /label= Repeat_8
 FT 169..173
 FT /label= Repeat_9
 FT 174..178
 FT /label= Repeat_10
 FT 179..183
 FT /label= Repeat_11
 FT 184..188
 FT /label= Repeat_12
 FT 187..192
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region
 FT 189..193

FT Region /label= Repeat_13
FT 190..197
FT /note= "Antigenic peptide, specifically claimed in
FT claim 15"
FT Region 194..198
FT /label= Repeat_14
FT Region 199..203
FT /label= Repeat_15
FT Region 202..215
FT /note= "Antigenic peptide, specifically claimed in
FT claim 15"
FT Region 204..208
FT /label= Repeat_16
XX US2002081701-A1.
XX 27-JUN-2002.
XX 03-AUG-2001; 2001US-0922480.
XX 17-MAR-1999; 99US-124820P.
XX 17-MAR-2000; 2000US-0527345.
XX (PADLE/) ADLER D A.
XX (SHEP/) SHEPPARD P O.
XX Adler DA, Sheppard PO;
XX WPI; 2002-635468/68.
XX N-PSDB; ABS52633, ABS52634.
XX Novel secreted salivary protein, zsig63 and polynucleotide encoding it
XX useful for treating microbial infections, inflammatory conditions,
XX dental caries and lung infections associated with cystic fibrosis
XX Claim 10; Page 28; 33pp; English.
XX The present invention relates to a new secreted salivary protein, zsig63.
XX The invention is useful for detecting in a test sample, the presence of
XX an antagonist or agonist of zsig63 protein activity. The invention is
XX also useful as an immunogen for producing an antibody to zsig63
XX polypeptide. zsig63-cytokine fusion proteins or antibody-cytokine fusion
XX protein are useful for enhancing in vivo killing of target tissues.
XX Pharmaceutical composition comprising purified zsig63 polypeptide are
XX useful in the treatment of conditions associated with pathological
XX microbes, including bacterial, fungal and viral infections. High
XX expression of zsig63 in salivary gland suggests that anti-microbial
XX polypeptides are useful for treatment of dental caries (tooth decay),
XX periodontal disease, thrush and gastrointestinal disease. Other
XX applications can be used in urinary tract infections, vaginal infections,
XX prevention of infection in skin and other epithelial wounds. The
XX polypeptides can be used to establish normal microflora and protect
XX against pathogenic colonisation and invasion. The invention is useful
XX when pro-inflammatory activity is desired. Applications for
XX such pro-inflammatory activity include the treatment of chronic tissue
XX damage, particularly in areas having a limited or damaged vascular system
XX e.g., damage in extremities associated with diabetes. Antagonists to
XX zsig63 polypeptides may be useful as anti-inflammatory agents. The
XX invention is useful for the treatment of patients having incompetent
XX immune system, such as AIDS (acquired immunodeficiency syndrome) patients
XX or individuals that have undergone chemotherapy, radiation treatment. The
XX invention is also useful for the treatment of lung infections associated
XX with cystic fibrosis and its agonists or antagonists are useful for
XX aiding digestion. The present amino acid sequence represents the human
XX secreted salivary protein zsig63 of the invention. This sequence is
XX encoded by the human zsig63 gene located on chromosome 4.
SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.5e-85;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLIIMACTVCVAFARRRPFFIGSDNDGDGHLPLPSLNIPYGINLPPPLYRRPVNTVP 60
DB 1 MLLIIMACTVCVAFARRRPFFIGSDNDGDGHLPLPSLNIPYGINLPPPLYRRPVNTVP 60
QY 61 SYRGNTYDTGCLPSYPMILTSRGPYVYHNGFPLATOLNVPLPPRGPFVPSRFSFA 120
DB 61 SYRGNTYDTGCLPSYPMILTSRGPYVYHNGFPLATOLNVPLPPRGPFVPSRFSFA 120
QY 121 AAAPAAPPIAAEPAAAPLTLATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
DB 121 AAAPAAPPIAAEPAAAPLTLATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
QY 181 EAPVGEPAEPSPAEPAATKAPAAEPHPSPGLEQANO 219
DB 181 EAPVGEPAEPSPAEPAATKAPAAEPHPSPGLEQANO 219

RESULT 4
ID AAU74536 standard; Protein; 219 AA.
AC AAU74536;
DT 23-APR-2002 (first entry)
XX Human zsig63 polypeptide.
XX Human; zsig63; chromosome 4q12-q413; salivary protein; antimicrobial;
XX microbial infection; tooth decay; periodontal disease; thrush; emphysema;
XX gastrointestinal disease; urinary tract infection; vaginal infection;
XX skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
XX acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
XX chronic bronchitis; gene therapy; protein therapy.
OS Homo sapiens.
XX US6331413-B1.
XX 18-DEC-2001.
XX 17-MAR-2000; 2000US-0527345.
XX 17-MAR-1999; 99US-124820P.
XX (ZYMO) ZYMOGENETICS INC.
XX Adler DA, Sheppard PO;
XX WPI; 2002-096707/13.
XX N-PSDB; AAS20591.
PT Polynucleotides encoding salivary proteins useful as anti-microbial
XX agents -
XX Claim 1; Column 49-52; 29pp; English.
XX The invention relates to a polynucleotide derived from the 4q12-q413
XX region of human chromosome 4 and encoding a zsig63 polypeptide, a
XX secreted salivary protein with anti-microbial activity. Due to their
XX microbial activity, the sequences can be used in the study of microbial
XX infections, e.g. for recombinant production of anti-microbial proteins.
XX The sequences can be used in the treatment of tooth decay, periodontal
XX disease, thrush, gastrointestinal disease, urinary tract infections,
XX vaginal infections, skin infections, epithelial wounds, chronic tissue
XX damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung
XX infections, sarcoidosis, emphysema and chronic bronchitis. This sequence
XX represents human zsig63.
SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.5e-85;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible][illegible]

DR N-PSDB; ABX93594.
XX Novel isolated zsig63 polypeptide, member of the adhesin family, useful
PT for treating dental caries, periodontal disease, thrush,
PT gastrointestinal disease, urinary tract infections, vaginal infections,
PT skin infections
XX
XX Claim 10; Page 27-28; 32pp; English.
XX
XX The invention relates to an isolated zsig63 polypeptide comprising at
CC least 90% identity to an amino acid sequence which comprises domain 1 of
CC zsig63, domain 2, domain 3, mature zsig63 and full length zsig63. Also
CC included are the polynucleotide encoding zsig63, a zsig63 expression
CC vector, a cultured cell comprising the vector and expressing the protein,
CC a DNA encoding a fusion protein (comprising amino acids 1-15, 16-37,
CC 38-126, 127-219 or 16-219 of zsig63 and an additional protein), using a
CC zsig63 reporter gene construct to identify zsig63 agonists, and
CC producing an anti-zsig63 antibody using zsig63 immunogenic peptides,
CC zsig63 is useful for detecting in a test sample, the presence of
CC antagonist of zsig63 protein activity. Zsig63 has antimicrobial
CC activity and since exhibits high expression in salivary gland, can be
CC used for treating dental caries, periodontal disease, thrush, and
CC gastrointestinal disease, urinary tract infections, vaginal infections,
CC skin infections and other epithelial wounds. The polypeptides can be
CC used to establish normal microflora and protect against pathogenic
CC colonization and invasion. Zsig63 can also be used for providing
CC pro-inflammatory activity for treating chronic, tissue damage
CC particularly in areas having limited or damaged vascular system, e.g.
CC in diabetes, and for treating immunocompromised AIDS patients or in
CC individuals that have undergone chemotherapy, radiation treatment, for
CC treating lung infections e.g. in cystic fibrosis. Detection of zsig63
CC polypeptide at relatively high levels in the trachea may indicate that
CC such polypeptides may serve as a marker of lung dysfunction. Zsig63 is
CC also useful in diagnosing conditions associated with salivary gland or
CC lung dysfunction including salivary gland carcinoma, pneumocystis carinii
CC infection, emphysema, chronic bronchitis, prostate dysfunctions such
CC as prostatic adenocarcinoma, aiding digestion, and as components of
CC defined cell culture media and may be used to replace serum that is
CC commonly used in culture. The DNA is useful in gene therapy applications
CC to increase or inhibit zsig63 activity, and for detecting abnormalities
CC on human chromosome 4 (e.g. 4q12-4q13), associated with dentinogenesis
CC imperfecta, and dentin dysplasia type II). Zsig63 is an adhesin family
CC member. The present sequence represents human zsig63.
XX
XX
SQ Sequence 219 AA;
Query Match 100.0%; Score 1178; DB 24; Length 219;
Best Local Similarity 100.0%; Pred. No. 2, 5e-85;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLLMAGTCVAPARKRRPPIGDDNDGHLPSLNIPIGIRNLPPPLYRPNVTP 60
DB 1 MKLLMAGTCVAPARKRRPPIGDDNDGHLPSLNIPIGIRNLPPPLYRPNVTP 60
QY 61 SYPGNTYTDGLSPYWIITSPGFPVYHIRGFPPLATQNLVPLPRGPPFPSPRFFSA 120
DB 61 SYPGNTYTDGLSPYWIITSPGFPVYHIRGFPPLATQNLVPLPRGPPFPSPRFFSA 120
QY 121 AAPAAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAPAAEAVAAEPAA 180
DB 121 AAPAAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAPAAEAVAAEPAA 180
QY 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219
DB 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219

RESULT 6
ID AAM30653 standard; Protein; 219 AA.
XX
AC AAM30653;
XX

DT 12-APR-1999 (first entry)
XX
XX DE Human secreted protein clone cp16 1 protein.
XX
XX KW Human; secreted protein; nutritional activity; cytokine; vaccine;
KW cell proliferation; differentiation; immune stimulation; suppression;
KW haematopoiesis regulation; tissue growth; actinin; inhibitor; chemotactic;
KW chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
KW tumour invasion suppression; tumour inhibition.
XX
XX OS Homo sapiens.
XX
XX PN WO9901466-A1.
XX
XX PD 14-JAN-1999.
XX
XX PF 01-JUL-1998; 98WO-US13813.
XX
XX PR 27-OCT-1997; 97US-0958304.
XX
XX PR 02-JUL-1997; 97US-0887195.
XX
XX PA (GENV) GENETICS INST INC.
XX
XX PI Agostino M, Jacobs K, Lavallee BR, McCoy JM, Racie LA;
PI Spaulding V, Treacy M;
XX
XX DR MPI; 1999-105994/09.
XX
XX DR N-PSDB; AAV80740.
XX
XX PT New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult testes, adult brain, foetal kidney,
PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
PT potential vaccines
XX
XX PS Claim 24; Page 71-72; 107pp; English.
XX
XX The present sequence represents a human secreted protein from clone
CC cp16 1, deposited as ATCC 98482. Human secreted protein clone
CC polynucleotides and proteins are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Suggested
CC activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccine) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, actinin/inhibitor activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy.
XX
XX
SQ Sequence 219 AA;
Query Match 99.5%; Score 1172; DB 20; Length 219;
Best Local Similarity 99.1%; Pred. No. 7, 4e-85;
Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLLMAGTCVAPARKRRPPIGDDNDGHLPSLNIPIGIRNLPPPLYRPNVTP 60
DB 1 MKLLMAGTCVAPARKRRPPIGDDNDGHLPSLNIPIGIRNLPPPLYRPNVTP 60
QY 61 SYPGNTYTDGLSPYWIITSPGFPVYHIRGFPPLATQNLVPLPRGPPFPSPRFFSA 120
DB 61 SYPGNTYTDGLSPYWIITSPGFPVYHIRGFPPLATQNLVPLPRGPPFPSPRFFSA 120
QY 121 AAPAAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAPAAEAVAAEPAA 180
DB 121 AAPAAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAPAAEAVAAEPAA 180
QY 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219
DB 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219

RESULT 7
AA19472
ID AA19472 standard; Protein; 221 AA.
XX
AC AA19472;
XX
DT 14-JUL-1999 (first entry)
XX
DE Amino acid sequence of a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testis disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS.
XX
OS Homo sapiens.
XX
PN W09922243-A1.
XX
PD 06-MAY-1999.
XX
PE 23-OCT-1998; 98WO-US22376.
XX
PR 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.
PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063097.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kaye H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
DR WPI: 1999-303069/25.
DR N-PSDB; AAX61352.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX
PS Claim 11; Page 401-402; 546pp; English.
XX
XX The specification describes cDNA sequences (AAX61322-X61470) encoding
XX human secreted proteins (AA19442-Y19590). The polynucleotides and their
XX corresponding secreted polypeptides are useful for preventing, treating
XX or ameliorating medical conditions, e.g. by protein or gene therapy.
XX Pathological conditions can also be diagnosed by determining the amount
XX of the polypeptides in a sample or by determining the presence of
XX mutations in the polynucleotides. Specific uses are described for each
XX of the polynucleotides, based on which tissues they are most highly
XX expressed in, and include developing products for the diagnosis or
XX treatment of cancer, tumours, neurodegenerative disorders, developmental
XX abnormalities and fetal deficiencies, blood disorders, leukemias,
XX diseases of the immune system, autoimmune diseases, hepatic and renal
XX diseases, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
XX and cognitive disorders, schizophrenia, prostate diseases, obesity,

CC disorders involving osteoclasts such as osteoporosis, arthritis or
CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
CC disorders, infections and AIDS. The polypeptides are also useful for
CC identifying their binding partners.
XX
SQ Sequence 221 AA;
XX
Query Match 94.5%; Score 1113.5; DB 20; Length 221;
Best Local Similarity 95.0%; Pred. No. 3.1e-80;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;
XX
QY 1 MKLLMAGYICVAPAKRRPPTIGEDNDGHLHPELNIPIYGRNLPPLYYRPVNTVP 60
DB 1 MKLLMAGYICVAPAKRRPPTIGEDNDGHLHPELNIPIYGRNLPPLYYRPVNTVP 60
QY 61 SYGNTTDTGDLGSPYMLTSPGPPVYHTRGPPLAQLVNPLPPRGPPVPPSPFSA 120
DB 61 SYGNTTDTGDLGSPYMLTSPGPPVYHTRGPPLAQLVNPLPPRGPPVPPSPFSA 120
QY 121 AAAPAAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAAP-VGAEPAAEPVAAEPA 179
DB 121 AAAPAAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVAAEPA 180
QY 180 AEAAPGVPAEAPBSPAPATKAPADBPBPSLEQANQ 219
DB 181 AEAAPGVPAEAPBSPAPATKAPADBPBPSLEQANQ 220
XX
RESULT 8
ABG21919
ID ABG21919 standard; Protein; 325 AA.
XX
XX ABG21919;
AC
DT 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #21910.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX W0200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAS86106.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 20; SEQ ID NO 52276; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 325 AA;

Query Match 15.6%; Score 183.5; DB 22; Length 325;
 Best Local Similarity 34.3%; Pred. No. 9.5e-07;
 Matches 57; Conservative 8; Mismatches 78; Indels 23; Gaps 6;

QY 48 PPPLYRPPVNTVSYGNTYTDGLSPYEWILSPGEPYVYHRCPLATQNVPLPR 107
 DB 45 PCPSRYCPQMTPPPP-----PAPPPALPPPP-----APAPPLPPAPSP- 87
 QY 108 GPFVPPSPSPFSAAPAPPAAPPAAPLTATPVAAEPAGAPVAAEPAAEPVGA 167
 DB 88 --PHAPPPQALPPPPALPPPP--ASPPLPPALSPPLPAPSPAPSPAPSPAPPPAP 143
 QY 168 PAAPAPVAAEPAAEPVGAEPAAEPSPAPAPATAPAPAPAPSP 212
 DB 144 PPSPPAPPPSPAPSPAPPLPPAPAPSPAPPPAPPPAPPPAPPP-PPSP 188

RESULT 9

AA040797
 ID AAR40797 standard; Protein; 207 AA.

AC AAR40797;
 DT 25-MAR-2003 (updated)
 DT 08-FEB-1994 (first entry)
 DE Sequence of a 10-amino acid repetitive sequence of an antigenic
 DE peptide domain encoded by an insert of clone Tcd.
 XX
 XX Epitope repeat; antigen; Chagas disease; TCD.
 XX
 OS Trypanosoma cruzi.
 OS
 XX

Key Location/Qualifiers
 FT Misc-difference 42
 FT /label= Degeneracy in repeat unit
 FT /note= "see also Aas 88,108,158,162"

W09316199-A1.

PD 19-AUG-1993.

PF 11-FEB-1993; 93WO-US01231.

PR 14-FEB-1992; 92US-0836642.

PA (REBD/) REED S G.

PT Reed SG;

DR WPI; 1993-272900/34.

DR N-PSDB; AA040797.

PT Antigenic peptide domain of trypanosoma cruzi - used to diagnose
 PT or immunise against infection, and screen blood supplies
 XX

PS Disclosure; Figure 1; 16pp; English.

XX Clone Tcd encodes a 10-amino acid repetitive sequence present
 CC in 20.5 copies with minor degeneracies present in 5 posns.
 CC The predicted mol. wt. of recombinant unglycosylated Tcd antigen
 CC is 36.3KD. Analysis of DNA from seven geographically diverse
 CC T. cruzi isolates indicates that Tcd gene sequence was conserved.
 CC A 636 bp fragment of clone Tcd was used to probe DNA from several
 CC other protozoan parasites of humans. However, it did not hybridise
 CC to T. brucei, Leishmania chagasi, L. amazonensis, L. donovani
 CC or T. rangeli.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC

SO Sequence 207 AA;

Query Match 15.3%; Score 180; DB 14; Length 207;
 Best Local Similarity 38.4%; Pred. No. 1.1e-06;
 Matches 43; Conservative 11; Mismatches 46; Indels 12; Gaps 2;

QY 114 PPRFSAAPAPAP-----PIAAEPAAAPLTATPVAAEPAGAPVAAEPAAEPV 164
 DB 85 PAESKSAEBKPAEPKSAEBKPAEPKSAEBKPAEPKSAEBKPAEPKSAEBK 144
 QY 165 GAEPAAEPVAAEPAAEPVGAEPAAEPSPAPAPATAPAPAPAPSP 213
 DB 145 PAEPKSAEBKPAEPKSAEBKPAEPKSAEBKPAEPKSAEBKPAEPKSAEBK 196

RESULT 10

AAW06915
 ID AAW06915 standard; Protein; 207 AA.

AC AAW06915;
 DT 02-APR-1997 (first entry)
 DT T. cruzi Tcd antigen.
 DE Tcd; diagnosis; infection; antigen; Chagas disease.
 XX
 XX Trypanosoma cruzi.
 OS

Key Location/Qualifiers
 FT Region 6..20
 FT /label= Antigenic epitope
 FT /note= "Claim 3, page 38"

W09629605-A2.

PD 26-SEP-1996.

PF 12-MAR-1996; 96WO-US03380.

PR 14-MAR-1995; 95US-0403379.

PA (CORI-) CORIXA CORP.

PT Reed SG;

DR WPI; 1996-485445/48.

DR N-PSDB; AAT46149.

PT Detecting and preventing T. cruzi infection - using polypeptide(s)
 PT or antibodies contg. or reactive with antigen epitope(s) of T. cruzi
 PT proteins

PS Disclosure; Fig 4; 59pp; English.

CC The Tcd antigen (AAW06915) of Trypanosoma cruzi includes an
 CC antigenic epitope (see also AAW06917) that can be utilized, pref.
 CC with other T. cruzi epitopes (see also AAW06914-19), in immunoassays
 CC for T. cruzi infection (Chagas' disease) and in vaccine compns.
 XX

SQ Sequence 207 AA;

Query Match 15.2%; Score 179; DB 17; Length 207;
 Best Local Similarity 41.0%; Pred. No. 1.3e-06;
 Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAAPAPPIAAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAEP 173

DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63

QY 174 VAAEPAAEPVGVPAEPSP-----AEPATAPPAEPHPSPS 213

DB 64 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 11

ID AAY32850 standard; Protein; 207 AA.

XX AAY32850;

DT 29-OCT-1999 (first entry)

DE TCD protein sequence.

XX TCD repeat sequence; 35kD T. cruzi homologue; immune system disorder;

XX eukaryotic ribosomal protein L19E; infection; detection; diagnosis;

XX Chagas' disease; Leishmania infection.

OS Trypanosoma cruzi.

XX US5942403-A.

XX 24-AUG-1999.

PF 15-SEP-1997; 97US-0929414.

PR 15-SEP-1997; 97US-0929414.

PR 14-MAR-1995; 95US-0403379.

XX (CORI-) CORIXA CORP.

XX Houghton R, Reed SG, Skeiky YAW;

XX WPI; 1999-517419/43.

XX N-PSDB; AAZ10979.

XX Trypanosoma cruzi antigenic homologue of eukaryotic ribosomal

XX protein L19E, useful for screening assays to detect T. cruzi

XX infection

XX Example 3; Fig 4; 26pp; English.

XX This sequence is the Trypanosoma cruzi Tcd protein. Antigenic fragments

XX of Tcd and fragments of the 35 kD Trypanosoma cruzi homologue (Tcd) of

XX the eukaryotic ribosomal protein L19E can be used in the method of the

XX invention for detecting T. cruzi infection in a biological sample, which

XX comprises: (a) contacting a biological sample with a first polypeptide

XX comprising the sequence shown in AAY32850; (b) contacting the biological

XX sample with a second polypeptide comprising the Tcd sequence shown in

XX AAY32850 or AAY32841; and (c) detecting the presence of antibodies that

XX bind to at least one of the polypeptides, indicating T. cruzi infection.

XX The methods and polypeptides may be used to detect T. cruzi (which

XX causes Chagas' disease and a variety of immune system disorders) and/or

XX Leishmania infection in individuals and blood supplies. The compounds

XX and methods may also be used to protect against T. cruzi infection. The

XX N-terminal region of Tcd (residues 1-136) cross-reacts with

XX anti-Leishmania antibodies and can be specifically excluded from the

XX antigenic polypeptide to avoid such cross-reactivity.

XX Sequence 207 AA;

SQ Query Match 15.2%; Score 179; DB 20; Length 207;

Best Local Similarity 41.0%; Pred. No. 1.3e-06;
 Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAAPAPPIAAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAEP 173

DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63

QY 174 VAAEPAAEPVGVPAEPSP-----AEPATAPPAEPHPSPS 213

DB 64 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 12

ID AAY23305 standard; Peptide; 207 AA.

XX AAY23305;

DT 31-AUG-1999 (first entry)

DE TCD antigen of Trypanosoma cruzi.

XX Trypanosoma cruzi epitope; Trypanosoma cruzi infection; antigen;

XX vaccine; Chagas' disease; TCD antigen.

OS Trypanosoma cruzi.

XX WO9931246-A1.

XX 24-JUN-1999.

PF 04-DEC-1998; 98WO-US28871.

PR 18-DEC-1997; 97US-0993674.

XX (CORI-) CORIXA CORP.

XX Houghton RL, Lodes WJ, McNeill PD, Reed SG, Skeiky YAW;

XX Smith JM;

XX WPI; 1999-405035/34.

XX New isolated Trypanosoma cruzi epitopes

XX Disclosure; Page 83-84; 103pp; English.

XX The specification describes new Trypanosoma cruzi epitopes. A method

XX for detecting Trypanosoma cruzi infection in a biological sample

XX comprises contacting the sample with a polypeptide comprising

XX an epitope of a TC antigen, or a variant of the antigen that differs

XX only in conservative substitutions and/or modifications and detecting

XX the presence of antibodies that bind to the polypeptide in the sample,

XX thereby detecting TC infection. The TC polypeptides can be used in

XX vaccines for inducing protective immunity against Chagas' disease in

XX a patient. The polypeptides and antibodies can also be used for detecting

XX TC infection. The present sequence represent a TCD antigen, which

XX can be used in the above assay to improve sensitivity.

SQ Sequence 207 AA;

Query Match 15.2%; Score 179; DB 20; Length 207;

Best Local Similarity 41.0%; Pred. No. 1.3e-06;

Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAAPAPPIAAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAEP 173

DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63

QY 174 VAAEPAAEPVGVPAEPSP-----AEPATAPPAEPHPSPS 213

DB 64 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 903 AA;
 SQ

Query Match 14.9%; Score 175.5; DB 23; Length 903;
 Best Local Similarity 32.3%; Pred. No. 1.2e-05;
 Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;

QY 35 HPSLNIPIYGRNLP---PPLYRPNVTVPSPYSGNTYTDGLPSYPMILTSPPGPPYYHNR 91
 DB 579 HKQVNVNNVTSSPTSNPVTTKPVTT-----KPVTTTKP-----VTTTKP----- 622
 QY 92 GFLPLATQLNVPPPLPRGPFVPPSPRFSAAAPAA-APPFAEPAAAPLTATFVAAPAA 150
 DB 623 ----VTIINQPSVKP-----AAAKPAPAKPVAAKPVATKTATVRP----- 658
 QY 151 GAPVAAEPAAEA-PVGAEPAAEA-PVAAEPAAEA-PVGEPAABEPSPAEPAATAKPAABEPH 209
 DB 659 --PVAVKPATAAKPVAAKPAAVRPPAA--AAAKPVATKEVPBPQAAKPAATKPAATTKPM 714
 QY 210 PPSPLE 215
 DB 715 VKMSRE 720

RESULT 15
 AAM40303
 ID AAM40303 standard; Protein; 3177 AA.
 XX
 AC AAM40303;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 3448.
 XX
 KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO20015312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.

XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J,
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AA159459.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 6; SEQ ID NO 3448; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX
 SQ Sequence 3177 AA;
 XX

Query Match 14.9%; Score 175.5; DB 22; Length 3177;
 Best Local Similarity 32.3%; Pred. No. 4.4e-05;
 Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;

QY 35 HPSLNIPIYGRNLP---PPLYRPNVTVPSPYSGNTYTDGLPSYPMILTSPPGPPYYHNR 91
 DB 2853 HKQVNVNNVTSSPTSNPVTTKPVTT-----KPVTTTKP-----VTTTKP----- 2896
 QY 92 GFLPLATQLNVPPPLPRGPFVPPSPRFSAAAPAA-APPFAEPAAAPLTATFVAAPAA 150
 DB 2897 ----VTIINQPSVKP-----AAAKPAPAKPVAAKPVATKTATVRP----- 2932
 QY 151 GAPVAAEPAAEA-PVGAEPAAEA-PVAAEPAAEA-PVGEPAABEPSPAEPAATAKPAABEPH 209
 DB 2933 --PVAVKPATAAKPVAAKPAAVRPPAA--AAAKPVATKEVPBPQAAKPAATKPAATTKPM 2988
 QY 210 PPSPLE 215
 DB 2989 VKMSRE 2994

Search completed: January 15, 2004, 15:13:23
 Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:10:34 ; Search time 35 Seconds
(without alignments)
1614.672 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 1178
Sequence: 1 MKLLMAGTICVAFARRRF.....TAKPAAPHPSPSLQANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	359.5	30.5	344	11	Q9D5J3	Q9D5J3 mus musculus
2	224.5	19.1	3179	12	Q8V2A4	Q8V2A4 human Herpe
3	221	18.8	1100	2	Q9X6J9	Q9X6J9 klebsiella
4	214.5	18.2	1480	10	Q9LIE8	Q9LIE8 arabidopsis
5	214	18.2	3105	12	Q8U2I8	Q8U2I8 cercopithec
6	206.5	17.5	607	5	Q17112	Q17112 babesia bov
7	204.5	17.4	1070	2	Q8GFS8	Q8GFS8 citrobacter
8	204	17.3	796	5	Q96579	Q96579 trypanosoma
9	199	16.9	391	5	Q26892	Q26892 trypanosoma
10	194.5	16.5	722	5	Q26893	Q26893 trypanosoma
11	193	16.4	3534	12	Q39266	Q39266 equine herp
12	188.5	16.0	1108	2	Q49542	Q49542 mycoplasma
13	188	16.0	839	16	Q9RX57	Q9RX57 deinococcus
14	186.5	15.8	395	16	Q9RTV4	Q9RTV4 deinococcus
15	186.5	15.8	616	11	Q99X31	Q99X31 mus musculus
16	183	15.5	744	10	O65375	O65375 arabidopsis

17	182	15.4	386	10	Q9FPQ5	Q9FPQ5 chlamydomon
18	181	15.4	928	12	Q9IMX9	Q9IMX9 cercopithec
19	181	15.4	2657	11	Q88493	Q88493 mus musculus
20	178.5	15.2	243	10	Q9XIV1	Q9XIV1 cucumis sat
21	177.5	15.1	598	16	Q8VKN7	Q8VKN7 mycobacteri
22	177	15.0	1315	10	Q9SEPW0	Q9SEPW0 zea mays (m
23	176.5	15.0	3084	12	Q8U2I1	Q8U2I1 pseudorabie
24	176	14.9	959	16	Q8XRH0	Q8XRH0 ralestonia s
25	175	14.9	857	16	Q98IK2	Q98IK2 rhizobium 1
26	174.5	14.8	979	4	Q8N421	Q8N421 homo sapien
27	174.5	14.8	1611	3	Q42854	Q42854 schizosach
28	174	14.8	236	10	Q9LM00	Q9LM00 pinus taeda
29	174	14.8	503	12	Q39779	Q39779 equine herp
30	173.5	14.7	786	10	Q48809	Q48809 arabidopsis
31	173	14.7	1174	4	Q94854	Q94854 homo sapien
32	171.5	14.6	438	16	Q9A2W5	Q9A2W5 cauliobacter
33	171.5	14.6	616	4	Q9H6K5	Q9H6K5 homo sapien
34	171	14.5	544	12	Q89392	Q89392 paramedum
35	171	14.5	1340	16	Q9L1H8	Q9L1H8 streptomyce
36	170.5	14.5	601	5	Q17113	Q17113 babesia bov
37	170	14.4	295	2	Q32850	Q32850 mycobacteri
38	170	14.4	428	10	Q94108	Q94108 zea mays (m
39	170	14.4	548	16	Q06404	Q06404 mycobacteri
40	170	14.4	842	5	Q9VGC9	Q9VGC9 drosophiila
41	170	14.4	864	5	Q95045	Q95045 drosophiila
42	170	14.4	864	5	Q9VGC8	Q9VGC8 drosophiila
43	170	14.4	1354	11	Q9EPW8	Q9EPW8 mus musculu
44	169.5	14.4	802	11	P70433	P70433 mus musculu
45	169	14.3	312	5	Q90NNT	Q90NNT drosophiila

ALIGNMENTS

RESULT 1	ID	Q9D5J3	PRELIMINARY;	PRT;	344 AA.
AC	Q9D5J3	Q9D5J3	Q9D5J3		
DT	01-JUN-2001	(T-REMBLrel. 17, Created)			
DT	01-JUN-2001	(T-REMBLrel. 17, Last sequence update)			
DT	01-JUN-2001	(T-REMBLrel. 17, Last annotation update)			
DE	4930432K09R1k	protein.			
GN	4930432K09R1k				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxId=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Testis;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Tani Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Glasi C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schirimi L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Guernstich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Liyone P., Marchionni L., Mashima J., Mazzarelli J., Monbets P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,				
RA	Hayashizaki Y.;				
RT	Functional annotation of a full-length mouse cDNA collection.;				
RL	Nature 409:685-690(2001).				
DR	EMBL; AK015291; BAB29782.1;				
DR	MGI; MGI:1921029; 4930432K09R1k.				

QY 130 AAEPAAPLTAATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPVAAEPVGA 189
DB 498 PGPPOAATTPGPPOAATTPGPPOAATTPGPPOAATTPGPPOAATTPGPPOAATTP 557
QY 190 AEEPSPAE--PATAKPAEPHPSPS 213
DB 558 PGPPOATSHAPOLPRAASAPAPQPTPT 583

RESULT 6

017112 PRELIMINARY; PRT; 607 AA.
AC 017112.
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
OS Babesia boydii.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
OX NCBI_TaxID=5865;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93341567; PubMed=8341317;
RA Dalrymple B.P., Peters J.M., Goodger B.V., Bushnell G.R.,
Walshbuhl D.J., Wright I.G.;
RT "Cloning and characterization of cDNA clones encoding two Babesia
boydii proteins with homologous amino- and carboxy-terminal domains";
RL Mol. Biochem. Parasitol. 59:181-189 (1993).
DR EMBL; M3126; AAA02753.1; -
SQ SEQUENCE 607 AA; 67129 MW; 2E06EC7DF843D732 CRC64;

Query Match 17.5%; Score 206.5; DB 5; Length 607;
Best Local Similarity 46.5%; Pred. No. 7.3e-07;
Matches 46; Conservative 7; Mismatches 35; Indels 11; Gaps 2;
QY 128 PTAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPVAAEPVGA 187
DB 329 PVEBEPVAAEPVVEBEPVVEBEPVVEBEPVVEBEPVVEBEPVVEBEPVVEBEPVVEB 388
QY 188 PAABEP-----SPAEPATAKPA---APEHPSPSLE 215
DB 389 PVABEPVABEPVCEPABETPAEPKAPETPAEPKAPETPAEE 427

RESULT 7

08GFS8 PRELIMINARY; PRT; 1070 AA.
AC 08GFS8;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE DNA primase.
GN FRI.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Bacteriia; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RA Goleblewski M., Zienkiewicz M., Adamczyk M., Kern-Zdanowicz I.,
RA Ceglowski P.;
RT "Complete nucleotide sequence of highly transmissible plasmid pCTX-
M3";
RT Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF550415; AAN87680.1; -
KW Plasmid.
SQ SEQUENCE 1070 AA; 117583 MW; F161DCB3137B500 CRC64;

Query Match 17.4%; Score 204.5; DB 2; Length 1070;
Best Local Similarity 30.9%; Pred. No. 1.7e-06;
Matches 69; Conservative 16; Mismatches 81; Indels 57; Gaps 7;

QY 9 IVCVAFARKRRFP-----FIGEDD--NDD-----GHPILSLINTPYGIRN 46
DB 298 MBEVAGRLKDTFEPSTHYFLDNDIYKQDENTGLEKATEAALTLGHVLPFNSP----- 352
QY 47 LPPPLYRFPVNTVDSYPGNTYTDGLSPYPIILTSPPGPVY-YHIRGFPLATOLNVPLP 105
DB 353 -----KEGLTDYNDLHVSEGLEQRLQVEG--AINQNNRVDTM 388
QY 106 PRGPPFPSPSPFSGAAAAPAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPVGA 165
DB 389 PTDNPNITDVHSSSTDSA-----AVAAPEKAAPVASTPAAEPVGAAPVAAEPVGA 443
QY 166 AEPAAEPVAAEPAAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGA 208
DB 444 AAPVASAPAAEPVETAPVASTPAAEPVGAAPVAAEPVGAEPVGAEPVGAEPVGAEPVGA 486

RESULT 8

096579 PRELIMINARY; PRT; 796 AA.
AC 096579;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface antigen PHG8T#5 (Fragment).
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tulahuen;
RA Garcia G.A., Bontempi B., Bua J., Ruiz A.M.;
RT "Molecular characterization of a Trypanosoma cruzi clone recognized by
an anti-Schistosoma japonicum glutathione-S-transferase serum";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF091620; AAC98688.2; -
FT NON TER 1 1
SQ SEQUENCE 796 AA; 83983 MW; DD4DCC06622533D8 CRC64;

Query Match 17.3%; Score 204; DB 5; Length 796;
Best Local Similarity 30.5%; Pred. No. 1.4e-06;
Matches 62; Conservative 15; Mismatches 62; Indels 64; Gaps 7;
QY 47 LPPPLYRFPV-----NTVPSVP-----GNTYTQGLSPYPIILTS 82
DB 411 IPPP-ERKPVPAAPAAATSSSVEPANERVTTNTQPTVPSPATAGPQTD----- 456
QY 83 GFPYVYHIRGFPLATOLNVPLPGRGPFV-----PSRPSAAAAPAPPIAEP 133
DB 457 -----QTLNASSVSPSGAAPSKAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 502
QY 134 AAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVGAEPVGAEPVGAEPVGA 193
DB 503 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 562
QY 194 SPAEPATAKPAEP-----HPSPS 213
DB 563 KPAEPKSAEPKPAEPKSAEPKPT 585

RESULT 9

026892 PRELIMINARY; PRT; 391 AA.
AC 026892;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Surface antigen (Fragment).
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U50901; ABA7806.1; -.
 DR EMBL; AF012905; AAC32526.1; -.
 DR InterPro; IPR000719; Prot_kinase.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP, 1.
 SQ SEQUENCE 1108 AA; 124903 MW; 239CF62D61E11FE7 CRC64;

Query Match 16.0%; Score 188.5; DB 2; Length 1108;
 Best Local Similarity 51.2%; Pred. No. 2.4e-05;
 Matches 42; Conservative 8; Mismatches 27; Indels 5; Gaps 1;

QY 124 PAAPPIAEPAAAPLTATPVAAEPAAGAPVAAEPAEAPVGAEPAAEAPVAAEPAEAP 183
 DB 813 PAAKPEAKPVAAKPEFTTKVAAKPE-----AAKPEAKPVAAKPEAAKPVAAKPEAAK 867
 QY 184 VGEPAEAPSPAPPAATAPPA 205
 DB 868 VAAKPEAKPVAAKPEAAKPEVA 889

RESULT 13

ID Q9RX57 PRELIMINARY; PRT; 839 AA.
 AC Q9RX57;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein DR0458.
 GN DR0458.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.U., Lam P., McDonald L., Uitterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001904; AAF10038.1; -.
 DR TIGR; DR0458; -.
 DR InterPro; IPR002965; P rich extensn.
 DR PRINTS; PR01217; PRICEXTENSN.
 DR KX Hypothetical protein; Complete proteome.
 SQ SEQUENCE 839 AA; 79759 MW; 3B6CC2DCCFEFPADE CRC64;

Query Match 16.0%; Score 188; DB 16; Length 839;
 Best Local Similarity 33.0%; Pred. No. 2e-05;
 Matches 64; Conservative 14; Mismatches 92; Indels 24; Gaps 9;

QY 27 DNDGHPHLSLPIGIRLPPPLYRPPVTPSPIGNTYDTGSPSYMITSP-GFP 85
 DB 140 DPTPAEFLKPP-----VQDTPPPVTPKPTPEPVTPKPAPEEVLQEPVACTPPVAKP 193
 QY 86 YVYHIGKFPPLAT-QLNVPLPPLPRGPFVPPSPRFSAAAAPPAEPIAAEPAAPLTATPV 144
 DB 194 PV-----PAPTQTPPTPPVQAPATRTTPPQ---AAKPTNAAGQAPPAATQAPPAQOTPT 244
 QY 145 AAEPAAGAPVAAEPAEAPVGAEPAAEAPVAAEPAEAPVGAEPAAEAPVAAEPAEAP 203
 DB 245 AQAAPATQTPATPAAPAAQAPAGAG--SPAAPAAQAAAPAGSVVPEATVDESSTPA--P 299
 QY 204 AAEPPHPSPLQEA 217
 DB 300 SAQTP-PTPTRETA 312

RESULT 14

ID Q9RV74 PRELIMINARY; PRT; 395 AA.
 AC Q9RV74;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein DR0938.
 GN DR0938.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.U., Lam P., McDonald L., Uitterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001946; AAF10516.1; -.
 DR TIGR; DR0938; -.
 DR KX Hypothetical protein; Complete proteome.
 SQ SEQUENCE 395 AA; 40003 MW; 75FD963717DAAPFO CRC64;

Query Match 15.8%; Score 186.5; DB 16; Length 395;
 Best Local Similarity 35.0%; Pred. No. 1.2e-05;
 Matches 57; Conservative 11; Mismatches 64; Indels 31; Gaps 9;

QY 60 PSLP-GNTYDTGSLSPYMITLSPGFPYVYHIGFPLATQLNVPLPRGPFVPPSPRF 118
 DB 53 PAEPYGAQPTTSIVSGAVVADSPTQ---AAPAAQTAGKIRPA--PAAP----- 102
 QY 119 SAAAPPAAP-----IAEPAAPLTATPVAAEPAAGAPVAAEPAEAPVGAEPAA 170
 DB 103 ----APKIPPLPPLPRMWPAPPTPPRVEETTTAASPTQPPVQAAPATQPTPTQTPAA 158
 QY 171 EAPVAAEPAEAPVGEPPAA--EESPAA--EAT-AXGAAPPP 208
 DB 159 QAPATQPPATPAF---EPAAEPAAPATTBPATPTEPAAPAP 198

RESULT 15

ID Q99KJ1 PRELIMINARY; PRT; 616 AA.
 AC Q99KJ1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to procollagen, type VI, alpha 3 (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Straubeberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005491; AAN05491.1; -.
 DR HSSP; P12111; 1KMT.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR002965; P rich_extensn.
 DR InterPro; IPR002035; VWF_A.

DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00014; Kunitz_Bpt1; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR01217; RICHTEXTEN.
DR Prodom; PD00022; Kunitz_Bpt1; 1.
DR SMART; SMO0060; FN3; 1.
DR SMART; SMO0131; KU; 1.
DR SMART; SMO0327; VWA; 1.
DR PROSITE; PS00280; BPT1_KUNITZ_1; 1.
DR PROSITE; PS50279; BPT1_KUNITZ_2; 1.
DR PROSITE; PS50234; VWA; 1.
FW Collagen; Proclease inhibitor; Serine protease inhibitor
FT NON TER 1 1
SQ SEQUENCE 616 AA; 66408 MW; FBA6CAGC2BB53B3 CRC64;

Query Match	15.8%	Score 186.5;	DB 11;	Length 616;
Best Local Similarity	32.4%	Pred. No. 1.8e-05;		
Matches 67; Conservative	25;	Mismatches 72;	Indels 43;	Gaps 12

Qy	21	PFPGEDNDGDHLEHLSLNI	PGIRNLPEPLYYR	NPVSPGNTVYDGLSPY	PILT	80
Db	253	PVLAKDEPAKPAQARAPAKPAKSAKLV	---	QEVHQP	PAQTSV	PAKPKP
Qy	81	SPGFPPVYTHIRGPLATOLN	PEPLPE	RGEFPV	---	PSRPSAAAPPAEPIAAEPA
Db	304	APQPAA	-----	AKVP	AK	---
Qy	137	APLTATPV	-----	AAEP	AGAPVAAEPA	---
Db	349	QPPAPQPVLTKSAAVNPASANKPVAT	WTATATATAPLAAKPA	AAKPAATRP	PLAA	408
Qy	186	VEPAEPE	-----	SPAEP	ATAKPAEP	208
Db	409	VRPATPEAPRQAKPAATKPAITTP	---	435		

Search completed: January 15, 2004, 15:14:11
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:06:54 ; Search time 18 Seconds

(without alignment)
572.158 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178

Sequence: 1 MKLLIMACIVCAFARRRRP.....TAKPAPEPPSPSLQANQ 219

Scoring table: BLOSUM62

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	180	15.3	555	1	GPI_CHLRE
2	175	14.9	1054	1	IP2_STIAU
3	175	14.9	3149	1	TEGU_EBV
4	174	14.8	518	1	TPM4_DROME
5	170	14.4	3176	1	CA36_HUMAN
6	169.5	14.4	802	1	ENAH_MOUSE
7	169	14.3	299	1	RL22_DROME
8	169	14.3	865	1	CPN_DROME
9	167.5	14.2	450	1	CV1_PARDE
10	166	14.1	810	1	NFM_BOVIN
11	165.5	14.0	439	1	XP2_XENLA
12	162	13.8	5147	1	PCLO_HUMAN
13	160.5	13.6	992	1	EBN6_EBV
14	158	13.4	1274	1	ENAM_MOUSE
15	154	13.1	353	1	CCPA_ACEXY
16	154	13.1	4499	1	DYHA_CHLRE
17	152.5	12.9	5038	1	PCLO_MOUSE
18	152	12.9	721	1	YK82_MYCTU
19	151	12.8	88	1	HB2_MISCO
20	151	12.8	211	1	TUBB_SOLTU
21	150	12.7	352	1	ALGP_PSEAE
22	148.5	12.6	464	1	SA32_HUMAN
23	147	12.5	1386	1	ZAP3_MOUSE
24	146.5	12.4	477	1	MAZ_MOUSE
25	145.5	12.4	361	1	IF35_MOUSE
26	145.5	12.4	1565	1	PAC_STRNU
27	145	12.3	475	1	SA2_MOUSE
28	145	12.3	2167	1	SHK1_RAT
29	144.5	12.3	477	1	MAZ_HUMAN
30	142.5	12.1	1083	1	MAZ2_HUMAN
31	142.5	12.1	1083	1	TRD3_HUMAN
32	142	12.1	2161	1	SHK1_HUMAN
33	141.5	12.0	316	1	CDNC_HUMAN

34	141.5	12.0	465	1	FXD1_HUMAN	Q16776	homo sapien
35	140.5	11.9	852	1	WS14_HUMAN	Q9np71	homo sapien
36	140	11.9	3421	1	TEGU_HAYEB	P28955	equine hiep
37	140	11.9	5085	1	PCLO_RAT	Q91k66	rattus norv
38	139.5	11.8	397	1	SEPL_MOUSE	Q62170	mus musculu
39	139.5	11.8	1157	1	BBC1_YEAST	P47068	saccharomyc
40	138.5	11.8	206	1	FRP1_MEDTR	Q43564	medicago tr
41	138.5	11.8	1183	1	DRP1_RAT	P54258	rattus norv
42	138	11.7	376	1	FXL2_HUMAN	P58012	homo sapien
43	138	11.7	2517	1	NGR2_HUMAN	Q9y618	h nuclear r
44	137	11.6	236	1	PRP_MEDSA	Q40358	medicago sa
45	136.5	11.6	1003	1	MBDE_HUMAN	Q96dne	homo sapien

ALIGNMENTS

RESULT 1
GPI_CHLRE STANDARD; PRT; 555 AA.
AC Q9FP06; 003927;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Baktaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
ON NCBI_TaxID=3055;
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Moessner J.P., Waffenschmidt S., Kilz S., Drees J., Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in plant hydroxyproline-rich glycoproteins";
RT Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL, PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.B.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins";
RT Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -!- FUNCTION: Major component of the outer cell wall W6 (crystalline) layer.
CC -!- SUBUNIT: Associates with GP2 and GP3.
CC -!- PTM: N-glycosylated and O-glycosylated.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AF309494; AAC45420.1; -;
CC EMBL; M58496; AAA69706.1; ALT_SEQ.
CC Glycosylated; Q9FP06; -;
CC InterPro: IPR002965; P-rich extensn.
CC InterPro: IPR003882; P-rich extensn.
CC PRINTS; PR01217; PRICHEXTENS.
CC PRINTS; PR01218; PSTEXTENSIN.
CC Glycoprotein; Repeat; Signal.
CC SIGNAL.
CC CHAIN 1 29
CC DOMAIN 30 555
CC DOMAIN 40 339
CC DOMAIN 259 279
CC CARBOHYD 399 399
CC CARBOHYD 455 455
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	493	493	N-LINKED (GLCNAc...)	(POTENTIAL).
SQ	SEQUENCE	555 AA;	54219 KM;	6A854AB045502F5 CRC64;	
Query Match		15.3%;	Score 180,	DB 1;	Length 555;
Best Local Similarity		25.7%;	Pred. No.	0.00089;	
Matches		56;	Conservative	17;	Mismatches 83; Indels 62; Gaps 5
Oy	33 PLHPSLNPVIGINLPP-PLYRRPVNTVSYGNTYTDTGHSYFWILTSGPFYYHIR	91			
Dd	158 PVPPSPRPVPSPAPSPSPSPSPSPSPSPAPSPAPVPSP----	APPSP-----			207
Oy	92 GFPLATQLNVPPIPRKGFVPSPSRFFSAALPAAPPIAAEBAAPAPLTATPVAAEBAG	151			
Dd	208 -----APVPSPSAPSPSPS-----	PAPSPSPSAPSPSPAPSPVPVPSPAP	252		
Oy	152 APVAAEPAEA-				
Dd	253 SPAPSPKPAPAPPPSPSPSPSPSPSPFPALTMPMPSPPSPSPAPPTPTPSPSPSPS	312			
Oy	176 AEPAAAPVGVGEPAABESPAPAPATKKAPAPPHSPS	213			
Dd	313 PVPPSPAHPVPPSPAPSPAPSPAPSPAPPTPSPSPSPS	350			

ID	IP2_STIAU	STANDARD;	PRT;	1054 AA.
AC	IP2_STIAU			
AD	P55875;			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Translation initiation factor IP-2.			
GN	INPB.			
OS	Stigmatella aurantiaca.			
OC	Bacteria; Proteobacteria; Delta proteobacteria; Myxococcales;			
OC	Cyctobacteriineae; Cyctobacteraceae; Stigmatella.			
OX	NCBI_Taxid=41;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DM4;			
RX	MEDLINE=97234648; PubMed=9079922;			
RA	Bremond L., Laalami S., Derflard B., Cenatiempo Y.;			
RT	"Translation Initiation factor IP2 of the myxobacterium Stigmatella			
RT	aurantiaca: presence of a single species with an unusual N-terminal			
RT	sequence.";			
RL	J. Bacteriol. 179:2348-2355(1997).			
CC	-1- FUNCTION: One of the essential components for the initiation of			
CC	protein synthesis. Protects formylmethionyl-tRNA from spontaneous			
CC	hydrolysis and promotes its binding to the 30S ribosomal subunits.			
CC	Also involved in the hydrolysis of GTP during the formation of the			
CC	70S ribosomal complex (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-1- SIMILARITY: BELONGS TO THE IP-2 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X87940; CAA61162.1; -			
DR	PIR; T43226; T43226.			
DR	HASP; P1351; IELO.			
DR	HMAP; MF_00100; -1.			
DR	InterPro; IPR0004795; EF_GTPbind.			
DR	InterPro; IPR004161; EFTU_D2.			
DR	InterPro; IPR000178; IP2.			
DR	InterPro; IPR006847; IP2_N.			
DR	InterPro; IPR005225; Small_GTP.			
DR	Pfam; PF00009; GTP_EFTU; 1.			
DR	Pfam; PF03144; GTP_EFTU_D2; 2.			

DR	Pfam; PF04760; IF2_N; 2.	
DR	ProDom; PD186100; IF2; 1.	
DR	TIGRFAMs; TIGR00487; IF-2; 1.	
DR	TIGRFAMs; TIGR00231; small_GTP; 1.	
DR	PROSITE; PS01176; IF2; 1.	
KW	Initiation factor; Protein biosynthesis; GTP-binding.	
FT	DOMAIN 555 703	G-DOMAIN.
FT	NP_BIND 561 568	GTP (BY SIMILARITY) .
FT	NP_BIND 607 611	GTP (BY SIMILARITY) .
FT	NP_BIND 661 664	GTP (BY SIMILARITY) .
SO	SEQUENCE 1054 AA; 11323 MW; FFD519530B5D0669 CRC64;	

```

Query Match 1: 14.9%; Score 175; DB 1; Length 1054;
Best Local Similarity 38.3%; Pred. No. 0.0029;
Matches 49; Conservative 8; Mismatches 51; Indels 20; Gaps 4

QY 96 ATQTNVPLPRGPPFVPPSRFSSAAAAPAAPPIAA--EPAAAAPLQAT--PVAAPPAAG 151
   : : : : :
Db 98 ASDVSSPPSPD-----VHEAGGAERAAASERVFBAAAQVEPVAAPRAAASBPAAAPKAT 152

QY 152 APVAAEPAAAEAFVGAEPAAAEAFVAAAEPAAEAFVGVEPAAAEBSPP-----AEPRT 200
   : : : : :
Db 153 APVAPPEPVEBAEKAAAPAPAPPEPTEAPKTEAFVAAAAPTAAEAPTPAPRTVPVTSGRRAAS 212

QY 201 AKPAAPPE 208
   : : : : :
Db 213 CRGAAPLP 220

```

ID	TEGU EBV	STANDARD;	PRT;	3149 AA.
AC	P03186;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-APR-1993 (Rel. 25, Last annotation update)			
DE	Large tegument protein.			
GN	BpFL1.			
OS	Epstein-barr virus (strain 895.8) (Human herpesvirus 4).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Gammaherpesvirinae; Lymphocryptovirus.			
OK	NCBI_TaxID=10377;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84270667; PubMed=6087149;			
RA	Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,			
RA	Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Segun C.,			
RA	Tufnell P.S., Barrell B.G.;			
RT	"DNA sequence and expression of the B95.8 Epstein-Barr virus genome.";			
RL	Nature 310:207-211 (1984).			
CC	-1- FUNCTION: TEGUMENT PROTEIN.			
CC	-1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 US36,			
CC	EHV-1 24, EBV BpFL1, HVS-1 64, VZV 22, AND HCMV UL48.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/1sb-sib.ch).			
CC	or send an email to license@1sb-sib.ch .			
CC	-----			
DR	EMBL; V01555; CAA24839.1; -.			
DR	PIR; G93065; Q0858.			
DR	InterPro; IPR005928; Herpes_teg_N.			
DR	Pfam; PF04843; Herpes_teg_N.1.			
SQ	SEQUENCE 3149 AA; 337954 MW; 3DD0C576587313D8 CRC64;			
Query Match	Similarity	14.9%;	Score 175;	DB 1; Length 3149;
Matches 48;	Conservative 19;	Mismatches 40;	Indels 34;	Gaps 6;

```

Db      290 PPVWLPFFRRRIITPVYLRPLPFSSTFSDGFPFAARKYBPAAKTNSSPP--SSPASAAPASAPAS 347
Qy      146 AAPAAGAEVAAPAAEAPVGAEPAAEAAPVAAEPAAEAAPVGVEPAABEEPSAPATAKP-A-204
Db      348 AAPASAAAPASAAPASAAADVASAPASAPASASPFLPIPIGLHTGTGVAPSPTRPPASSCA 407
Qy      205 APE-----PHSPS 213
Db      408 APQTPTKRKKGLGXDSPHKKPT 428

RESULT 4
ID#     TPW4_DROME          STANDARD;   PRT;       518 AA.
AC      P49455; P49456; Q24425; Q24426;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Tropomyosin 1, isoforms 33/34 (Tropomyosin II).
GN      TM1 OR TM11.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
CX      NCBI_taxid=7227;
[1]
RN      SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RP      TISSUE=Embryo, and Pupae.
RX      MEDLINE=89127197; PubMed=2851721;
RA      Hanke P.D., Storci R.V.;
RT      "The Drosophila melanogaster tropomyosin II gene produces multiple
RT      proteins by use of alternative tissue-specific promoters and
RT      alternative splicing."
RL      Mol. Cell. Biol. 8:3591-3602(1988).
[2]
RP      SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC      STRAIN=Oregon-R; TISSUE=Pupa;
RX      MEDLINE=87064486; PubMed=3097506;
RA      Karlik C.C., Fyrberg E.A.;
RT      "Two Drosophila melanogaster tropomyosin genes: structural and
RT      functional aspects."
RL      Mol. Cell. Biol. 6:1965-1973(1986).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=5;
Comment=Additional isoforms seem to exist;
Name=33; Synonyms=9C;
IsoId=P49455-1; Sequence=Displayed;
Name=Muscle; Synonyms=9D;
IsoId=P06754-1; Sequence=External;
Name=Non-muscle; Synonyms=Cytoskeletal;
IsoId=P06754-2; Sequence=External;
Name=9A;
IsoId=P06754-3; Sequence=External;
Name=34; Synonyms=9B;
IsoId=P49455-2; Sequence=VSP 006623, VSP 006624, VSP 006625;
-1- TISSUE SPECIFICITY: Both isoforms are only expressed in indirect
flight muscles
-1- DEVELOPMENTAL STAGE: Both isoforms are expressed during pupal and
adult stages.
-1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
-1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
-----
Cc      This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc      the European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc      or send an email to license@isb-sib.ch).
OR      EMBL; X76208; CAA53800.1; -.

```

Query Match	Similarity	Score	DB 1	Length	DB 2	Indels	Gaps
Beat Local	30.2%	Pred. No. 0.0018					
Matches	77	Conservative	13	Mismatches	107	Indels	58
							7
17	KRRPFFIGEDNDGDHPLHPSLNIPIYIRULPPLIYRPIVNT						58
264	KERICWIGDSLDEAFVDLIGLGEFFPMWRPKPPTPKLPPTTBELLAAMEBAPAAALAA 323						
59	-----VPSYCGNTYTDGLPSYPIWLT--SDGPFVYHI--R 91						
324	AABAABAAABAAAAGADGADGAPAAAFGESEKAPAKGFTPTPKPTPPPPPPPPFVSDLPB 363						
92	GFPLATQULNPPLPFRPFVPVPSRFRFSAAAAP--AAPPI-AAEPALAAELTATPVAAEP 146						
384	GAIVPYKXNVEPPPPGSEPFVPAABEGGAAPAAEGAAPPEGAAPPEGAAPVPPADGAAP 443						
149	AAGAPVAABPAABAVGABPAABAPV-----AABPAAPVGVPEPAE-----EPSPAE 196						
444	ABGAAPAAABEGAAPPADGAAPPAABAAAAPADAAAAPAAABAAAPAAABAAAPAAABAA 503						
199	ATAKPAABEPHPS 213						

Db 504 ABAAPAAABGEAPPA 518

RESULT 5
 CAS6_HUMAN
 ID CA36_HUMAN STANDARD; PRT; 3176 AA.
 AC P12111; Q16501;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 15-JUN-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 3 (VI) chain precursor.
 GN COL6A3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Fibroblast;
 RX MEDLINE=90151612; PubMed=1699238;
 RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,
 Ghanville R., Mayer U., Mann K., Deutzmann R., Timpl R.;
 RT Mosaic structure of globular domains in the human type VI collagen
 alpha 3 chain: similarity to von Willebrand factor, fibronectin,
 RT actin, salivary proteins and apocytin type protease inhibitors.";
 RL EMBO J. 9:385-393(1990).
 RN [2]
 RP REVISIONS.
 RA Chu M.-L.;
 RN Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE OF 2038-2373 FROM N.A.
 RX MEDLINE=89066644; PubMed=3198591;
 RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,
 Timpl R.;
 RT "Amino acid sequence of the triple-helical domain of human collagen
 type VI.";
 RL J. Biol. Chem. 263:18601-18606(1988).
 RN [4]
 RP SEQUENCE OF 2092-2157 FROM N.A.
 RX MEDLINE=88029444; PubMed=3665927;
 RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,
 Hu-Chen C.-C., Bernard M.P., Timpl R.;
 RT "Characterization of three constituent chains of collagen type VI by
 peptide sequences and cDNA clones.";
 RL Eur. J. Biochem. 168:309-317(1987).
 RN [5]
 RP SEQUENCE OF 2092-2151 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88161046; PubMed=3348213;
 RA Weill D., Mettel M.-G., Passage B., van Cong N., Pribula-Conway D.,
 Mann K., Deutzmann R., Timpl R., Chu M.-L.;
 RT "Cloning and chromosomal localization of human genes encoding the
 three chains of type VI collagen.";
 RL Am. J. Hum. Genet. 42:435-445(1988).
 RN [6]
 RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=93054780; PubMed=1339440;
 RA Zanussi S., Doliama R., Segat D., Bonaldo P., Colombatti A.;
 RT "The human type I collagen gene. mRNA and protein variants of the
 alpha 3 chain generated by alternative splicing of an additional 5-end
 exon.";
 RL J. Biol. Chem. 267:24082-24089(1992).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.
 RX MEDLINE=95182468; PubMed=753217;
 RA Arnoux B., Merigean K., Saludjian P., Norris F., Norris K., Bjoern S.,
 Olsen O., Petersen L., Ducruix A.;
 RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of
 human type VI collagen.";
 RL J. Mol. Biol. 246:609-617(1995).
 RN [8]

RP STRUCTURE BY NMR OF 3102-3164.
 RX MEDLINE=96398604; PubMed=8805527;
 RA Zweckstetter M., Cziech M., Mayer U., Chu M.-L., Zinth W., Timpl R.,
 Holak T.A.;
 RT "Structure and multiple conformations of the Kunitz-type domain from
 human type VI collagen alpha3 (VI) chain in solution.";
 RL Structure 4:195-209(1996).
 RN [9]
 RP STRUCTURE BY NMR OF 3107-3164.
 RX MEDLINE=97410331; PubMed=925624;
 RA Soerensen M.D., Bjoern S., Norris K., Olsen O., Petersen L.,
 James T.L., Led J.J.;
 RT "Solution structure and backbone dynamics of the human alpha3-chain
 type VI collagen C-terminal Kunitz domain.";
 RL Biochemistry 36:10439-10450(1997).
 RN [10]
 RP DISEASE.
 RX MEDLINE=21987636; PubMed=11992252;
 RA Demir E., Sabatelli P., Allamand V., Ferreira A., Moghadazadeh B.,
 Makrelouf M., Topaloglu H., Echeine B., Merlini L., Guicheney P.;
 RT "Mutations in COL6A3 cause severe and mild phenotypes of Ullrich
 congenital muscular dystrophy.";
 RL Am. J. Hum. Genet. 70:1446-1458(2002).
 RN [11]
 RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.
 RX MEDLINE=98204804; PubMed=9536084;
 RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,
 Stajich J.M., Viles K., Vance J.M., Chu M.-L., Speer M.C.;
 RT "Missense mutation in a von Willebrand factor type A domain of the
 alpha 3 (VI) collagen gene (COL6A3) in a family with Bethlem
 myopathy.";
 RL Hum. Mol. Genet. 7:807-812(1998).
 CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (VI),
 CC -1- ALPHA 2 (VI), AND ALPHA 3 (VI).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named Isoforms=2;
 CC Name=1;
 CC IsoId=P12111-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P12111-2; Sequence=VSP 001172;
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- DISEASE: Defects in COL6A3 are a cause of Bethlem myopathy (BM)
 CC [MIM:158810]. BM is a rare autosomal dominant proximal myopathy
 CC characterized by early childhood onset (complete penetrance by the
 CC age of 5) and joint contractures most frequently affecting the
 CC elbows and ankles.
 CC -1- DISEASE: Defects in COL6A3 are a cause of Ullrich congenital
 CC muscular dystrophy (UCMD) [MIM:254090]; also known as Ullrich
 CC scleroatonic muscular dystrophy, an autosomal recessive congenital
 CC myopathy. UCMD is characterized by muscle weakness and multiple
 CC joint contractures, generally noted at birth or early infancy. The
 CC clinical course is more severe than in Bethlem myopathy.
 CC -1- SIMILARITY: Contains 1 BPT/Kunitz inhibitor domain.
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -1- SIMILARITY: Contains 12 VMPA domains.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, X52022; CA36267.1; -;
 DR EMBL, X06196; CA329557.1; -;
 DR EMBL, M20778; -; NOT ANNOTATED_CDS.
 DR EMBL, M27449; AA52057.1; -;
 DR EMBL, S69432; AAB34261.1; -;
 DR PIR, AS9140; CGH03A.
 DR PDB, 1KNT; 01-NOV-94.

DR PDB; 2KNT; 15-MAY-97.
 DR PDB; 1KUN; 12-NOV-97.
 DR PDB; 1KTH; 28-AUG-02.
 DR Genew; HGNC:2213; COL6A3.
 DR MIM; 120250; -.
 DR MIM; 158810; -.
 DR MIM; 254050; -.
 DR GO; GO:0005589; C:collagen type VI; TAS.
 DR GO; GO:0005202; P:collagen; TAS.
 DR GO; GO:0007517; P:muscle development; TAS.
 DR InterPro; IPR000867; Collagen.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR002035; VWFA_A.
 DR Pfam; PF00014; Collagen; 5.
 DR Pfam; PF00092; Kunitz_BPTI; 1.
 DR PRINTS; PR00759; BASICPRASE.
 DR ProDom; PD000007; C1g_helix; 2.
 DR ProDom; PD000222; Kunitz_BPTI; 1.
 DR SMART; SM00060; FN3; 1.
 DR SMART; SM00131; KU; 1.
 DR SMART; SM00327; VWFA; 12.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
 DR PROSITE; PS50234; VWFA; 12.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;
 KW Signal; 3D-structure; Disease mutation; Polymorphism;
 KW Alternative splicing.
 KM
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 3176 COLLAGEN ALPHA 3 (VI) CHAIN.
 FT DOMAIN 26 2038 NONHELICAL REGION.
 FT DOMAIN 2039 2375 TRIPLE-HELICAL REGION.
 FT DOMAIN 2376 3176 NONHELICAL REGION.
 FT 39 213 VWFA 1.
 FT DOMAIN 242 419 VWFA 2.
 FT 445 620 VWFA 3.
 FT DOMAIN 639 816 VWFA 4.
 FT DOMAIN 837 1009 VWFA 5.
 FT 1029 1205 VWFA 6.
 FT 1233 1404 VWFA 7.
 FT 1436 1609 VWFA 8.
 FT DOMAIN 1639 1812 VWFA 9.
 FT 1838 2024 VWFA 10.
 FT DOMAIN 2402 2581 VWFA 11.
 FT 2619 2815 VWFA 12.
 FT DOMAIN 2987 3076 FIBRONECTIN TYPE-III.
 FT DOMAIN 3107 3176 BPTI/KUNITZ INHIBITOR.
 FT SITE 2040 2042 CELL ATTACHMENT SITE.
 FT SITE 2136 2138 CELL ATTACHMENT SITE.
 FT SITE 2148 2150 CELL ATTACHMENT SITE.
 FT SITE 2154 2156 CELL ATTACHMENT SITE.
 FT SITE 2370 2372 CELL ATTACHMENT SITE.
 FT ACT_SITE 3121 3122 REACTIVE BOND.
 FT DISULFID 3111 3161
 FT DISULFID 3120 3144
 FT DISULFID 3136 3157
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT
 Query Match 14.4% Score 170; DB 1; Length 3176;
 Best Local Similarity 31.8%; Pred. No. 0.014;
 Matches 57; Conservative 16; Mismatches 56; Indels 50; Gaps 9;

Qy 151 GAPIVAEPAAEA-EVGAEPAAEAFAVGEVPAAPAEPSAPAEPAATKAPAP 208
 Db 2933 --PVAIVPAPATAKAVAAKPAAPVPPA---AAKAVATKPEVPRPQAKPAATKPAATTTP 2986
 RESULT 6
 ENAH_MOUSE STANDARD; PRT; 802 AA.
 AC Q03173; P70430; P70431; P70432; P70433;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Enabled protein homolog (NPC derived proline-rich protein 1) (NDPP-1).
 GN ENAH OR MENA OR NDPE1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=93041923; PubMed=1420303;
 RA Sazuka T., Tomooka Y., Kachju S., Ikawa Y., Noda M., Kumar S.;
 RT "Identification of a developmentally regulated gene in the mouse
 RT central nervous system which encodes a novel proline rich protein.";
 RL Biochim. Biophys. Acta 1132:240-248(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3, 4 AND 5), FUNCTION, AND SUBCELLULAR
 RP LOCATION.
 RC TISSUE=Brain;
 RX MEDLINE=97015079; PubMed=8861907;
 RA Gertler F.B., Niebuh K., Reinhard M., Wehland J., Soriano P.;
 RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
 RT control of microfilament dynamics.";
 RL Cell 87:227-239(1996).
 RN [3]
 RP FUNCTION, AND SUBUNIT.
 RX MEDLINE=99166867; PubMed=1069337;
 RA Lanier L.M., Gates M.A., Witke W., Menzies A.S., Wehman A.M.,
 RA Macklis J.D., Kwiatkowski D., Soriano P., Gertler F.B.;
 RT "Mena is required for neurulation and commissure formation.";
 RL Neuron 22:313-325 (1999).
 CC -1- FUNCTION: May be involved in microfilament assembly and cell
 CC motility. Induces the formation of F-actin rich outgrowths in
 CC fibroblasts. Required for neurulation and commissure formation.
 CC -1- SUBUNIT: Binds profilin.
 CC -1- SUBCELLULAR LOCATION: Localized to focal adhesions and, to a
 CC lesser extent, leading edges and stress fibers.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Name=5; Synonym=Mena+++;
 CC Name=5; Synonym=Mena+++;
 CC Name=1; IsoId=Q03173-1; Sequence=Displayed;
 CC Name=2; IsoId=Q03173-2; Sequence=VSP_007255, VSP_007256;
 CC Name=3; Synonym=Mena; IsoId=Q03173-3; Sequence=VSP_007259, VSP_007260;
 CC Name=4; Synonym=Mena+++; IsoId=Q03173-4; Sequence=VSP_007259;
 CC Name=5; Synonym=Mena+++; IsoId=Q03173-5; Sequence=VSP_007257, VSP_007258;
 CC -1- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
 CC LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
 CC -1- SIMILARITY: Contains 1 WH1 domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; D10727; BAA01570.1; --


```

DR FlyBase; FBgn015288; RPL22.
DR InterPro; IPR002671; Ribosomal_L22e.
DR Pfam; PF01776; Ribosomal_L22e; 1.
DR ProDom; PD007306; Ribosomal_L22e; 1.
KW Ribosomal protein.
FT DOMAIN 24 31 POLY-ALA.
FT DOMAIN 46 50 POLY-ALA.
FT DOMAIN 65 70 POLY-ALA.
FT DOMAIN 93 98 POLY-ALA.
FT DOMAIN 103 112 POLY-ALA.
FT DOMAIN 136 152 POLY-ALA.
FT DOMAIN 185 188 POLY-LYS.
FT DOMAIN 299 ASP/GLU-RICH (HIGHLY ACIDIC).
SQ SEQUENCE 239 AA; 30610 MW; 46A9005610E4EB0 CRC64;

Query Match 14.3%; Score 169; DB 1; Length 299;
Best Local Similarity 44.1%; Pred. No. 0.0022;
Matches 49; Conservative 6; Mismatches 34; Indels 22; Gaps 5;

Qy 121 AAAPAPPIAPPAAPLTAAT-----PVAAPPAAGAPVAAPPAAPVGAAPPAAP 173
Db 65 AAAAARPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAP 124

Qy 174 VAAAPPA-----AAAPVGPAAEPPSPAPPAATKPAAPPPH-----PSPS 213
Db 125 AAAPPAKKAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAP 171

RESULT 8
CPN_DROME STANDARD; PRT; 865 AA.
AC 002910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Calpuctin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin U.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calpuctin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calpuctin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -!- FUNCTION: Might function as a calcium-sequestering "sponge" to
regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
of Ca(2+) per mole of protein.
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCELLI.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; L02111; AAA28405.1; -
CC EMBL; L05080; AAA28420.1; -
CC PIR; A47282; A47282.
CC PIR; A47283; A47283.
DR FlyBase; FBgn010218; Cpn.
DR GO; GO:0005509; F:calcium ion binding activity; IDA.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> AP (IN REF. 2).
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417B0B0E7CFE CRC64;

Query Match 14.3%; Score 169; DB 1; Length 865;
Best Local Similarity 29.5%; Pred. No. 0.0053;
Matches 70; Conservative 14; Mismatches 75; Indels 78; Gaps 11;

Qy 55 PVATVPSYFGNTYTDGLSEYPMILTSFGPPYYTHNGFPLAT---QLNPPLEPPRGPF 111
Db 42 PIAVTPAAPPTTASV---QPAVTIPAPAPPIAASGTPVAVAPVVAAPTPPAASPV 97

Qy 112 VPP---SRPFAAAPAPAPPIAEP-----AAAPLTAT-PVAAB---PAA----- 150
Db 98 STPVAAQIIVAAVAPAPVVAATPTTPVQIPVAAPIAIPVVAASPTPAATPVISPV 157

Qy 151 -----GAPVAAP-----AAAPVGA-----PAAAPVA 175
Db 158 IASPPVANTTVVAPVAAVPAVPAVLAAPVAPVAPVATPAPVAPVAPV 217

Qy 176 ABAAPAPV-----GEPAAEPSPAPPTATCAPAPPPHS---PSIEQA 217
Db 218 TIECVAPLIPVSVVATKPLAAAPVVAAPATETTPVVAAPVAPVAVETA 274

RESULT 9
CY1_PARDP STANDARD; PRT; 450 AA.
AC P13627;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome c1 precursor.
GN PPTC.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxId=266;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007612; PubMed=2820981;
RA Kuriowski B., Ludwig B.;
RT "The genes of the Paracoccus denitrificans bc1 complex. Nucleotide
sequence and homologues between bacterial and mitochondrial
subunits.";
RL J. Biol. Chem. 262:13805-13811(1987).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS. C1 FUNCTIONS AS AN ELECTRON DONOR TO
CYTOCHROME C.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -----

```



```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - -
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation - -
CC CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi-sb.ch/announce/ or send an email to license@ebi-sb.ch).
```

```

OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
SIMILARITY).
-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR DR EMBL; AF091342; AAC6357.1; -.
DR DR InterPro; IPR001664; IP.
DR DR InterPro; IPR002957; Keratin_I.
DR DR Pfam; PF00038; Filament; 1.
DR DR PRINTS; PRO1248; TYPE1KERATIN.
DR DR PROSITE; PS00226; IF; 1.
KM Intermediate filament; Coiled coil; Neutrone; Phosphorylation.
FT FT NON TER 1
FT FT 1 296 ROD.
FT FT DOMAIN 297 810 TAIL.
FT FT DOMAIN <1 33 COIL 1A.
FT FT DOMAIN 21 33 LINKER 1.
FT FT DOMAIN 34 132 COIL 1B.
FT FT DOMAIN 133 149 LINKER 12.
FT FT DOMAIN 150 171 COIL 2A.
FT FT DOMAIN 172 175 LINKER 2.
FT FT 176 296 COIL 2B.
FT FT 503 582 8 X 13 AA TANDEM REPEATS.
SQ SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;

Query March 14.1%; Score 166; DB 1; Length 810;
Best Local Similarity 35.1%; Pred. No. 0.0073;
Matches 34; Conservative 12; Mismatches 51; Indels 0; Gaps 0;

QY 112 VPSRPFSSAAAPAAPPIAPAAAADLTTPVAABEPAAGAPVAAEPAAEPVGAEPALAE 171
Db 484 VAPKEELAAEAKEKPEPKAKSPVAKSPPTTKSPAKSPSEAKSPKASPTAKSPVAK 543

QY 172 APVAAEPAAEPVGVPEPAABEESPBAEPATKAAPAEP 208
Db 544 SPTAKSPSEAKSPKASPTAKSPAKSPAPKSP 580

RESULT 11
XP2_XENLA STANDARD; PRT; 439 AA.
ID XP2_XENLA
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SBP-2003 (Rel. 42, Last annotation update)
DS Skin secretory protein xp2 precursor (ABEG protein).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).
RX MEDLINE=92332564; PubMed=16928230.
RA Hauser F., Roeben C., Hoffmann W.,
RT "xp2, a new member of the P-domain peptide family of potential growth
RT factors, is synthesized in xenopus laevis skin.";
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).
RC TISSUE=Skin;
EX MEDLINE=90127399; PubMed=2298293;
RA Gnachi M., Berger H., Thallhammer J., Kreil G.;
RT "dermal glands of Xenopus laevis contain a polypeptide with a highly

```

h1y

CC Isoleid=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC Isoleid=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y19188; CAB60727.1; -;
 CC EMBL; AC004903; AAD20936.1; -;
 CC EMBL; AC004886; AAD21789.1; -;
 CC EMBL; AB011131; BAA25485.1; -;
 CC EMBL; BC001304; AAB01304.1; -;
 CC EMBL; AC004082; AAB97937.1; -;
 CC PIR; T00634; T00634.
 CC HSP; P04410; IAA25.
 CC Genew; HGNC:13406; PCLO.
 CC MIM; 604918; -;
 CC GO; GO:0005856; C:cytoskeleton; NMS.
 CC GO; GO:0045202; C:synaptic junction; ISS.
 CC GO; GO:0005509; F:calcium ion binding activity; ISS.
 CC GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
 CC GO; GO:0005522; F:profilin binding activity; ISS.
 CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR001565; Synaptotagmin.
 CC PRINTS; PRO0360; C2DOMAIN.
 CC PRINTS; PRO0339; SYNAPTOTAGMIN.
 CC SMART; SM00239; C2; 2.
 CC DR PROSITE; PS00499; C2 DOMAIN 1; 1.
 CC DR PROSITE; PS50004; C2 DOMAIN 2; 2.
 CC KW Calcium/phospholipid-binding; zinc; Metal-binding; zinc-finger;
 CC KW Repeat; Alternative splicing.
 CC FT 1
 CC FT 1
 CC FT 400 465
 CC FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 CC FT P-A-K-P-Q-P-Q-Q-P-X.
 CC FT C4-TYPE (POTENTIAL).
 CC FT C4-TYPE (POTENTIAL).
 CC FT POLY-PRO.
 CC FT PDZ.
 CC FT C2 DOMAIN 1.
 CC FT C2 DOMAIN 2.
 CC FT S -> SGNGIIRIVGKEIPGHSGEIGAYIAKILPGSGAS
 CC FT QTEGKLMVE (in isoform 2).
 CC FT /Frid=VSP_003923.
 CC FT K -> KPTDGTGVSHPIRGEID (in isoform 2).
 CC FT /Frid=VSP_003924.
 CC FT G -> GQVAVVQNAS (in isoform 2).
 CC FT /Frid=VSP_003925.
 CC FT TARKS -> SKRRK (in isoform 2).
 CC FT /Frid=VSP_003926.
 CC FT Missing (in isoform 2).
 CC FT /Frid=VSP_003927.
 CC FT /Frid=VSP_003927.
 CC FT SEQUENCE 5147 AA; 563537 MW; CDSD64990496CDDC CRC64;
 CC Query Match 13.8%; Score 162; DB 1; Length 5147;
 CC Best Local Similarity 23.9%; Pred. No. 0.057;
 CC Matches 60; Conservative 31; Mismatches 78; Indels 82; Gaps 10;
 CC QY 31 GHPIHSLNIPYGRILPPPLVYRPVNTVPSYGNITYTDG---LPSYMWILTSPPFY 87

DB 255 GESVYPSLPSF-----SKPPIQPTPGKPPAQPGHKSQPPAPKPPAQPSGLTK----- 304
 QY 88 YHNGFPLATQANV--PELPGRG-----PVPSPRRFSA----- 120
 DB 305 -----PLAQPGTYVYKPVVQPPGTTKPPAQPLGPAKPPAQQTSGEKSSSQPGKALAQ 358
 QY 121 -----AAAPAP-----PIAEPAAAPL-----TATPVAEPAPAPVAE 157
 DB 359 PGVGTTPAQGGPAKPPYQVGTGPKPLAQGQLGSPAPAPPTKTPAQTKPPSQPGSGTK 418
 QY 158 PAAPAPVAEPAPAPAPVAEPAPAPVAEPAPAPVAEPAPAPAT-----AKPA 205
 DB 419 PPPQPGPAKSPPOGSKTPSPSQPGSAKSAQPPSAKSAQFTKPVSGTGFGLQ 478
 QY 206 PEPHPSFLAQ 216
 DB 479 P-PTVSPSAKQ 488
 DB RESULT 13
 DB EBN6_EBV
 DB ID_EBN6_EBV STANDARD, PRT, 992 AA.
 AC P03204;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DB EBN6-6 nuclear protein (EBNA-3C) (EBNA-4B).
 GN BBRP3-BBRP4.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SOURCE: FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Watnall B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211 (1984).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=88155772; PubMed=2831394;
 RA Petri L., Sample J., Wang F., Kleff B.;
 RT "A fifth Epstein-Barr virus nuclear protein (EBNA3C) is expressed in
 RT latently infected growth-transformed lymphocytes";
 RL J. Virol. 62:1330-1338 (1988).
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=90266473; PubMed=2161150;
 RA Petri L., Sample C., Kleff B.;
 RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
 RT latent infection nuclear proteins";
 RL Virology 176:563-574 (1990).
 CC -1- FUNCTION: INVOLVED IN LATENT CYCLE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
 CC -1- SIMILARITY: SOME SIMILARITIES EXIST BETWEEN EBNA 4, 5, AND 6.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; VO1555; CA24859.1; -;
 CC DR Pfam; PF05009; EBNA-3; 1.
 CC KW Nuclear protein; Repeat.
 CC FT 74
 CC FT 80
 CC FT 551 610
 CC FT 741 779
 CC FT POLY-ARG.
 CC FT 10 X 5 AA TANDEM REPEATS.
 CC FT 3 X 13 AA TANDEM REPEATS.

SQ SEQUENCE 353 AA; 37396 MW; 287456CASE7FB6F CRC64;

Query Match 13.1%; Score 154; DB 1; Length 353;

Best Local Similarity 29.9%; Pred. No. 0.017; Mismatches 73; Indels 40; Gaps 7;

Matches 52; Conservative 9; Mismatches 73; Indels 40; Gaps 7;

```
QY 53 YRPVNTVPSYPGNTYTDGLPSYPWILTSBGPYVYHIRGFPLATQINVPPLPPRGFPFV 112
   |||
   : : : : :
DB 37 YRP-----FVDRSPDVTGVPE----AVERHFDQAEH----DTAVEBQVTPAP----- 75
   |||
   : : : : :
QY 113 PPSRFFSAAAPAPPIAAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEAAPVGAEPAAEA 172
   |||
   : : : : :
DB 76 -----QIIVAPPPPVDPDPPIVTEIAPFPV--VVSAPVTYBPPAAAVPAEPVQQA 127
   |||
   : : : : :
QY 173 PVAAEPAAAPVGVBPAAEPSPAPB-----ATAKPAPBPSPSLBOA 217
   |||
   : : : : :
DB 128 PVQAAFPVPAP--VPPIABQAPPAAPDPAASVPYANVAAPVPPDPAPVTPAPQA 179
```

Search completed: January 15, 2004, 15:12:28
Job time : 19 secs

A:Molecule type: DNA
 A:Residues: 1-395 <WHI>
 A:Cross-references: GB:AB001946; GB:AB000513; NID:g6458655; PIDN:AAFI0516.1; PID:g645866
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0938
 A:Map position: 1

Query Match 15.8%; Score 186.5; DB 2; Length 395;
 Best Local Similarity 35.0%; Pred. No. 0.0001;
 Matches 57; Conservative 11; Mismatches 64; Indels 31; Gaps 9;

QY 60 PSYV-GNTYTDGTPSYPMWLTSGFPYVYHKGFPATQLVNPLPRGFPVPSGRFF 118
 DB 53 PAEPVGAPOPTTISIVSGAVADGSPQT---AAPAOQTGAKIPPA--PAAP----- 102
 QY 119 SAAAPAAP-----IAEPAAAAPLTATPVAAPAGAPVAAAPAAAPVGAAPAA 170
 DB 103 ---APKIPPEVLEPRRMPAPTPPPRVEETTTAASPTQPPVTAQAPATQTPPTPPAA 158
 QY 171 EAPVAAEPAAAPVGVGPAA-EESPAA---EPAT-AKPAAP 208
 DB 159 QAPATQPPATPAP---EPAAAPAPATTPPATTPTEPAAP 198

RESULT 7

E86255

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C/Accession: E86255

R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, V.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: E86255

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-744 <STO>

A:Cross-references: GB:AB005172; NID:g3157926; PIDN:AACT17609.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 15.5%; Score 183; DB 2; Length 744;

Best Local Similarity 28.9%; Pred. No. 0.00031;

Matches 55; Conservative 14; Mismatches 93; Indels 28; Gaps 6;

QY 33 PLHPSLNIPIGIRNLPPPLLYRPNVTVPSYFGNTYDTGTPSY-----PMWLTSGFP 85
 DB 461 PPSPPSPPPYVSSPPPPYV---SSPPPPYVSSPPPPYVSSPPPPYVSSPPPPYVSSPPPP 516
 QY 86 YVYHIRFPLATQLVNPLPRGFPVPSRFSAAAAPAA--PPIAAAPAAAPLTATP 143
 DB 517 YVY-----SSPPPPPP--PSPPPPCPSSPPPPVYVYVQSPSPSPSPYVPP 562
 QY 144 VAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPVAAEPVGAEPVGAEPVGAEPVGAEP 203
 DB 563 VTGSPSPSPVYVPPVNSPPSPSPVYVPPVYVPPVYVPPVYVPPVYVPPVYVPPVYVPP 622
 QY 204 AAEPHPS 213
 DB 623 VTSP--PPPS 631

RESULT 8

A40215

Tcd antigen - Trypanosoma cruzi

C:Species: Trypanosoma cruzi

C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999

C/Accession: A40215

R/Burns Jr., J.M.; Shreffler, W.G.; Rosman, D.E.; Sleeth, P.R.; March, C.J.; Reed, S.G.

Proc. Natl. Acad. Sci. U.S.A. 89, 1239-1243, 1992

A>Title: Identification and synthesis of a major conserved antigenic epitope of Trypano

A:Reference number: A40215; MUID:92159024; PMID:1371355

A/Accession: A40215

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-207 <BUR>

A>Note: sequence extracted from NCBI backbone (NCBIN:82113, NCBI:P:82117)

C:Superfamily: histone H1

Query Match 15.2%; Score 179; DB 2; Length 207;

Best Local Similarity 41.0%; Pred. No. 0.00016;

Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSMAAPAPPIAEPAAAPLTATPVAAPAGAPVAAEPVGAEPVGAEPVGAEP 173
 DB 5 PAEPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 63
 QY 174 VAAEPAAEPVGVGPAAEPSP-----AEPATKAPAAEPHPS 213
 DB 64 KPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 108

RESULT 9

T4326

translation initiation factor IF-2 - Stigmatella aurantiaca

C:Species: Stigmatella aurantiaca

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Feb-2001

C/Accession: T4326

R/Bremund, L.; Ismail, S.; Derjard, B.; Ceniempo, Y.

U. Bacteriol. 179, 2348-2355, 1997

A>Title: Translation initiation factor IF2 of the myxobacterium Stigmatella aurantiaca:

A:Reference number: Z23352; MUID:97234648; PMID:9079922

A/Accession: T4326

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1054

A:Cross-references: EMBL:X87940; PIDN:CAA61162.1

A:Experimental source: strain DW4

C:Genetics:

A:Gene: infB

C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu hom

C/Keywords: nucleotide binding; P-loop; protein biosynthesis

F/555-664/Domain: translation elongation factor Tu homology <ETU>

F/561-568/Region: nucleotide-binding motif A (P-loop)

Query Match 14.9%; Score 175; DB 2; Length 1054;

Best Local Similarity 38.3%; Pred. No. 0.0013;

Matches 49; Conservative 8; Mismatches 51; Indels 20; Gaps 4;

QY 96 ATQANVPPLPPGFPVPSRFSAAAAPAPIAA--EPAAAAPLTAT--PVAAAPAG 151
 DB 98 ASDVSSPPSP-----VHSAAGAAASASRVPAALVQGEVAAAPAAASBPAAAPKAT 152
 QY 152 APVAAEPAAEPVGAEPAAEPVAAEPVAAEPVGAEPVGAEPVGAEPVGAEPVGAEP 200
 DB 153 APVAAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEPVGAEP 212
 QY 201 AKPAAP 208
 DB 213 CKGAAPLP 220

RESULT 10

Q08B8

Bp1F1 protein - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4, Epstein-Barr virus

C>Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999

C/Accession: G93065; A03747; S32993
 R/Banker: A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
 Moll. Biol. Med. 1, 21-45, 1993
 A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A/Reference number: A93065; MUID:85055713; PMID:6092825
 A/Accession: G93065
 A/Molecule type: DNA
 A/Residues: 1-3149 <BAN>
 A/Cross-references: EMBL:V01555; NID:959074; PIDN:CAA24839.1; PID:q1334853
 R/Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A/Reference number: A03794; MUID:84270667; PMID:6087149
 A/Contents: annotation; protein coding region
 C/Superfamily: human herpesvirus 4 BBLFL protein

Query Match 14.9%; Score 175; DB 1; Length 3149;
 Best Local Similarity 34.0%; Pred. No. 0.0039;
 Matches 48; Conservative 19; Mismatches 40; Indels 34; Gaps 6;

QY 105 PPRGPF-----VP-----PSRFSAAAPAA-----PPIAEPMAAALTTATPVA 145
 DB 290 PPRMLDPERRRIITPYNLRPLPSRSTSDSPPAATYSAKTNSPP--SSPASAPASAPAS 347
 QY 146 AEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAAEPVGAEPAAEPSPAPATAKP-A 204
 DB 348 AAPASAAAPASAAAPASAAAPASAPASAPSPPLFTPIPLGLHTPGVAPASTPPASGA 407
 QY 205 AP-----PHSPS 213
 DB 408 APQTPRRKKGLGKDSPHKKT 428

RESULT 11

T38236
 hypotetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C/Accession: T38236
 R/Murphy, L.; Harri, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1998
 A/Reference number: Z21780
 A/Accession: T38236
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1611 <MUR>
 A/Cross-references: EMBL:AL021813; PIDN:CAA16991.1; GSPDB:GN00066; SPDB:SPAC23A1.17
 A/Experimental source: strain 972h-; cosmid c23A1
 C/Genetics:
 A/Gene: SPDB:SPAC23A1.17
 A/Map position: 1

Query Match 14.8%; Score 174.5; DB 2; Length 1611;
 Best Local Similarity 32.5%; Pred. No. 0.0022;
 Matches 64; Conservative 18; Mismatches 84; Indels 31; Gaps 11;

QY 33 PLHPSINITYGIRNLPPLYRPNVTVSGNTYDTGLSPYWLTSPPGPPYVYHNG 92
 DB 1025 PL-PSADAF-----PIVPSATAPVPIPTGTTPYKSSGAPSPAPVPPAPS-----SEIFS 1075
 QY 93 PPIATQNLVPLP-PRGPFPPPSRFSAAAD-----AAPPIAEPMAAALTTATPVA 145
 DB 1076 ITPFS--GAPVPAPAPGIPVPPKP----SYAAPVPVPKPSVAVPPVPA-PGAGAPVPKPSVA 1129
 QY 146 AEPAA--AGA-----VAAEPAAEPVGAEPAAEPVAAEPAAEPVGAEPAAEPSPAP 197
 DB 1130 APVPVPKPSVAVPPVPAAPVPAAPVPAAPVPAAPVPAAPVPAAPVPAAPVPAAPVPAAPV 1188
 QY 198 PATAKPAEPHPSPL 214
 DB 1189 PVPVPSEAPVPKPSV 1205

RESULT 12

T01456
 extenion homolog F2401.18 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
 C/Accession: T01456
 R/Shim, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; C
 ecologists, A.; Ecker, J.R.
 submitted to the EMBL Data Library, January 1998
 A/Description: Genomic sequence for Arabidopsis thaliana BAC F2401.
 A/Reference number: Z14211
 A/Accession: T01456
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-786 <SHI>
 A/Cross-references: EMBL:AC003113; NID:92609438; PID:g2781362
 A/Experimental source: cultivar Columbia
 C/Genetics:
 A/Map position: 1
 A/Note: F2401.18

Query Match 14.7%; Score 173.5; DB 2; Length 786;
 Best Local Similarity 29.7%; Pred. No. 0.0013;
 Matches 62; Conservative 15; Mismatches 97; Indels 35; Gaps 9;

QY 33 PLHPSINITYGIRNLPPLYRPNVTVSGNTYDTGLSP--YFWLTSPPGPPYVYH 90
 DB 511 PPSPPPPPY-IXSPPPVAVNCP-PTQSPPPKKEQPSPREYV-----SPSPYYQYT 564
 QY 91 RGP-----LATQNLVPLPPRGF-----PVPSPFSFAAAPAAAPPIAEPMAA-- 137
 DB 565 SPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSP 624
 QY 138 PLTATPVAAPAAEPVGAEPAAEPVGA-----EPAAEPVAAEPAAEPV 184
 DB 625 PVTSPVTVQSPPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSP 684
 QY 185 GVEPAEPSPAPATAPAAEPSPSPS 213
 DB 685 YYPVPAKSPPPSPPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSPPPPPPTTAAQSP 712

RESULT 13

S52418
 GTP-binding regulatory protein Gs alpha-XL chain - rat
 N/Alternate names: G protein XL-alpha-8
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 02-Feb-2001
 C/Accession: S52418
 R/Kehlenbach, R.H.; Matthey, J.; Huttner, W.B.
 Nature 372, 804-809, 1994
 A/Title: XL-alpha-8 is a new type of G protein.
 A/Reference number: S52418; MUID:95089824; PMID:7997272
 A/Accession: S52418
 A/Molecule type: mRNA
 A/Residues: 1-846 <KEH>
 A/Cross-references: EMBL:X84047; NID:9642267; PIDN:CAA58866.1; PID:9642268
 R/Kehlenbach, R.H.; Matthey, J.; Huttner, W.B.
 Nature 375, 253, 1995
 A/Title: Correction: XLaiphae is a new type of G protein.
 A/Reference number: S58911
 A/Contents: annotation; assignment of start codon
 A/Note: experimental data from this paper suggest that the translation is initiated at
 C/Keywords: GTP binding; nucleotide binding; P-loop; signal transduction
 F/132-846/Product: GTP-binding regulatory protein Gs alpha-XL chain #status experiments
 F/499-506/Region: nucleotide-binding motif A (P-loop)
 F/744-747/Region: GTP-binding NKXD motif

Query Match 14.6%; Score 172.5; DB 2; Length 846;
 Best Local Similarity 33.0%; Pred. No. 0.0015;
 Matches 77; Conservative 15; Mismatches 88; Indels 53; Gaps 13;

QY 24 GEDD-----NDGHPHPSINITYGIRNLPP-----PPLYYRPNVTVPS-YFGNTYDT 70


```

Db      16 GVDDDTAVVMDSPPI-ASGPPPIEVSAGAPVXSEHAKRPIERQAAETGNSPISSTTAEEA 74
Qy      71 GLPSYPMILTSRQGFYVYHIRGFPLATQ-----LNVPLPRPGPFVPP 114
       75 KVPSSLERGESPTQPTWTHIKPAPVABSGTDSKADPSATTAVLIGPEVGVGVTMTPL 134
Db      115 SRFPSAAAAAPAPPIAEP-AAAAPLTATPVAA-----EPAAAPVAAEPAAEPVGAEP 169
       135 D--LPPASDEAGDVAAEPDGGTAP--ATPASEDNREPPAAA--AAEPAAEP--AAEP 186
Qy      170 AE--APVAAEPAAEA-----FVGEPPAAEPSPAPATAPAAEPHPSP 212
       187 AEPAAEPAAEPAAEAAPVDTAEASAGAVPDQEPAPAAAAAATPAPAPAAAP 239

```

RESULT 14

```

arylesterase-related protein (imported) - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87675
R.Nierman, W.C.; Feildbyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; WUID:21173698; PMID:11259647
A:Accession: G87675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <STO>
A:Cross-references: GB:AE005673; NID:g13425157; PIDN:AAK25403.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3441

```

```

Query Match      14.6%; Score 171.5; DB 2; Length 438;
Best Local Similarity 33.3%; Pred. No. 0.00094;
Matches 68; Conservative 14; Mismatches 59; Indels 63; Gaps 11;

Qy      64 GNTYYDTGILSPY---WLTSGRF-----PYVYHIRGFPLATQNLVPLPRPGPFVPP 114
Db      219 GKRTASADIKIYPNNNAHYLIDEPNTMKIIDDVIAWLRGKTQAPAAKVEBPAA----- 270
Qy      115 SRFPSAAAAAP-APPIAEPAAAAAPLTATPV--AAEPAA-----GAPVAAEP-----A 160
Db      271 ----PAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 323
Qy      161 EAPVG-----AEPAAEPV--AAEPAAEPVGEPPAAEPSPAP----- 197
Db      324 KAPVAKKAPKAKAAPAAAPAAAPKAAAPKAAAPKAAAPKAAAPKAAAPKAAAPKAAAPKAA 363
Qy      198 --PATKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219
Db      384 TAPAAKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPATK 407

```

RESULT 15

```

T17547
Proline-rich protein A57R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17547
R.Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17547
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-544 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAK96425.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:

```

```

A:Gene: A57R
Query Match      14.5%; Score 171; DB 2; Length 544;
Best Local Similarity 27.8%; Pred. No. 0.0012;
Matches 58; Conservative 19; Mismatches 106; Indels 26; Gaps 4;

```

```

Qy      2 KILLWACIVCAFAKRRP---PTGEDNDGDHPLHPLNIPLYGIRNLPPPLYRPNVT 58
Db      4 RVIFPLTLVTTAFAPABPPTTQIVTQTTVNSPKPSPKPSKPSKPSKPSKPSKPSKPS 63
Qy      59 VPSYGNVTTDTGLSPYPMILTSRQGFYVYHIRGFPLATQNLVPLPRPGPFVPPSRFP 118
Db      64 KPPSPKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPS 100
Qy      119 SAAAAPAPPIAEPAAAAAPLTATPVAAEPAAAPVAAEPAAEPVGAEPAAEPVAAEP 178
Db      101 PGPKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPS 160
Qy      179 AAAPVGVAAEPAAEPSPAPATAPAPAP 207
Db      161 PAPBPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 189

```

Search completed: January 15, 2004, 15:14:44
Job time : 23 secs

Db 181 EAPVGEPAEAEPSPAEPATAKPAABEPHPSLEQANQ 219

US-09-923-236-2
RESULT 2

Sequence 2, Application US/09923236
Patent No. US20020090677A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/923,236
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-923-236-2

Query Match 100.0%; Score 1178; DB 9; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.6e-77;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHPSINIPYGINLPPLPYRRPVNTVP 60

Db 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHPSINIPYGINLPPLPYRRPVNTVP 60

Qy 61 SYPGNTYTDGLPSYPMILTSRGPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120

Db 61 SYPGNTYTDGLPSYPMILTSRGPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120

Qy 121 AAAPAPPIAEPAAAPLTATPVAAEPAAAGPVAAPAAEPVGAEPAAEPVAAEPAA 180

Db 121 AAAPAPPIAEPAAAPLTATPVAAEPAAAGPVAAPAAEPVGAEPAAEPVAAEPAA 180

Qy 181 EAPVGEPAEAEPSPAEPATAKPAABEPHPSLEQANQ 219

Db 181 EAPVGEPAEAEPSPAEPATAKPAABEPHPSLEQANQ 219

RESULT 3
US-09-922-469-2

Sequence 2, Application US/09922469
Patent No. US20020173027A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/922,469
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-469-2

Query Match 100.0%; Score 1178; DB 10; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.6e-77;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHPSINIPYGINLPPLPYRRPVNTVP 60

Db 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHPSINIPYGINLPPLPYRRPVNTVP 60

Qy 61 SYPGNTYTDGLPSYPMILTSRGPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120

Db 61 SYPGNTYTDGLPSYPMILTSRGPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120

Qy 121 AAAPAPPIAEPAAAPLTATPVAAEPAAAGPVAAPAAEPVGAEPAAEPVAAEPAA 180

Db 121 AAAPAPPIAEPAAAPLTATPVAAEPAAAGPVAAPAAEPVGAEPAAEPVAAEPAA 180

Qy 181 EAPVGEPAEAEPSPAEPATAKPAABEPHPSLEQANQ 219

Db 181 EAPVGEPAEAEPSPAEPATAKPAABEPHPSLEQANQ 219

RESULT 4
US-09-746-783-106

Sequence 106, Application US/09746783
Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallee, Edward R.
Racie, Lisa A.
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.
Howes, Steven H.
Fechtel, Kim

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM

NUMBER OF SEQUENCES: 231

CORRESPONDENCE ADDRESSES:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/746,783

FILING DATE: 21-Dec-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mlaelincic, Debra J.

REGISTRATION NUMBER: 46,931

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:

LENGTH: 219 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 106:

US-09-746-783-106

Query Match 99.5%; Score 1172; DB 11; Length 219;
Best Local Similarity 99.1%; Pred. No. 7e-77;
Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHPSINIPYGINLPPLPYRRPVNTVP 60

Db 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHPSINIPYGINLPPLPYRRPVNTVP 60

Qy 61 SYPGNTYTDGLPSYPMILTSRGPYVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120

```

Db      61 SYPGNTYTDGLSPYMWILTSPPGFVYVHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
Qy      121 AAAPAPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAA 180
        121 AAAPAPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAA 180
        121 AAAPAPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAA 180
Qy      181 EAPVGEPAAEPSPAEPATAKPADEPHPSPSLEQANQ 219
        181 EAPVGEPAAEPSPAEPATAKPADEPHPSPSLEQANQ 219
Db      181 EAPVGEPAAEPSPAEPATAKPADEPHPSPSLEQANQ 219

```

RESULT 5

```

US-10-231-417-192
; Sequence 192, Application US/10231417
; Publication No. US20030176681A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1
; CURRENT APPLICATION NUMBER: US/10/231,417
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 192
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals stop translation
US-10-231-417-192

```

Query Match 94.5%; Score 1113.5; DB 12; Length 221;

Best Local Similarity 95.0%; Pred. No. 1.1e-72;

Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

```

Qy      1 MLLTLACTIVCVAFARRRRPPIGEDNDGPHLSLNPYGRMLPPPLYRPNVTV 60
        1 MLLTLACTIVCVAFARRRRPPIGEDNDGPHLSLNPYGRMLPPPLYRPNVTV 60
Db      1 MLLTLACTIVCVAFARRRRPPIGEDNDGPHLSLNPYGRMLPPPLYRPNVTV 60
Qy      61 SYPGNTYTDGLSPYMWILTSPPGFVYVHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
        61 SYPGNTYTDGLSPYMWILTSPPGFVYVHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
Db      61 SYPGNTYTDGLSPYMWILTSPPGFVYVHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
Qy      121 AAAPAPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAA 179
        121 AAAPAPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAA 179
Db      121 AAAPAPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAA 180
Qy      180 AEAAPVGEPAAEPSPAEPATAKPADEPHPSPSLEQANQ 219
        180 AEAAPVGEPAAEPSPAEPATAKPADEPHPSPSLEQANQ 219
Db      181 AEAAPVGEPAAEPSPAEPATAKPADEPHPSPSLEQANQ 220

```

RESULT 6

```

US-10-316-253-95
; Sequence 95, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M

```

```

; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 95
; LENGTH: 2657
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-316-253-95

```

Query Match 15.4%; Score 181; DB 12; Length 2657;

Best Local Similarity 29.1%; Pred. No. 0.00012;

Matches 57; Conservative 22; Mismatches 61; Indels 56; Gaps 8;

```

Qy      40 IYGRIRNLPP-----LYRPVNTVPSIPGNTYTDGLSPYMWILTSPPGFVYVHIRG 93
        40 IYGRIRNLPP-----LYRPVNTVPSIPGNTYTDGLSPYMWILTSPPGFVYVHIRG 93
Db      2239 VRFGRKQINPPTANSLSKVVTTIK--PSNSSKPTAL-----VNLTP----- 2280
Qy      94 PLATQLVNPPRPGPFVPSRFFSA 133
        94 PLATQLVNPPRPGPFVPSRFFSA 133
Db      2281 -----AKPAPAPAPAPQPVLAKEPAKPAQAPAPAPAPAPAPAPAPAPAPAPAP 2333
Qy      134 AAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAAEPAAEPAAEP 192
        134 AAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAAEPAAEPAAEP 192
Db      2334 ASVRPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 2389
Qy      193 PSPAPATKPADEPHPSPSLEQANQ 208
        193 PSPAPATKPADEPHPSPSLEQANQ 208
Db      2390 PMPAPVLTSAAVRP 2405

```

RESULT 7

```

US-10-264-049-2409
; Sequence 2409, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2409
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC_FEATURE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (15)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (795)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2409

```

Query Match 14.9%; Score 175.5; DB 12; Length 903;

Best Local Similarity 32.3%; Pred. No. 9.9e-05;

TYPE: PRT
ORGANISM: Homo sapiens
US-10-023-523-43

Query Match 14.7%; Score 173; DB 14; Length 538;
Best Local Similarity 36.8%; Pred. No. 8.9e-05;
Matches 49; Conservative 7; Mismatches 47; Indels 30; Gaps 5;

QY 102 PPLPFRGFVPPSRFPFAAAPPPIAAEPAAALTLTTPVAABPAAGAPVAABPAE 161
DB 111 PPAAPRPAAP-----AAAAAAPPPTPAPPPPPPAVAAAARARAAAATAAPSP 162
QY 162 APVGAEPAAE-APVAA-EPAAEPVGEV-----AAEP-----SPAEP 169
DB 163 GPAPGPRAPRAAPPAAPPAAPPAVAPBAPBAPRAAPPAVAAEPPLPPPPQAPAPQ 222
QY 200 TAKPAABEPHSP 212
DB 223 QQQPPPPQPPPP 235

RESULT 11

US-10-029-386-33686
Sequence 33686, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 33686
LENGTH: 980
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL137853.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.84
OTHER INFORMATION: SWISSPROT HIT: Q02910, EVALU 2.00e-03
US-10-029-386-33686

Query Match 14.7%; Score 173; DB 12; Length 980;
Best Local Similarity 31.1%; Pred. No. 0.00016;
Matches 68; Conservative 13; Mismatches 112; Indels 26; Gaps 9;

QY 6 MACIVCAAPAKRRPFPIGEDDNDGRLHPSLNI-----PYGIRNLPPPLYRRPVT 58
DB 531 WAAIAIV-----PTBEDGTREG-PVTPATVHAPEBDRAAVSVTPEEASPA 582
QY 59 VPSYPTNTYDTGLPSYPMWLTSPGFYVYHNGFPLATQLNVPPLPRGFPVP-PSRF 117
DB 583 VPTPEEPTSPAAPVT-PEEPTSPAAPVPTPEEPTSPAAPVPTPEEPTSPA 641
QY 118 FSAAPAAAPPIAAEPAAALTLTTPVAABPAAGAPVAABPAE 174
DB 642 TSPAAAPVPTPEEPTSPAAPVPTPEEPTSPAAPVPTPEEPTSPA 697
QY 175 AAEPAAPVGVPAABEPSPAPBAPATKAPABEPHSPS 213
DB 698 AAVPTPEEP--ASPAAVPTPEEPAFAPAVPTPEEAS 734

RESULT 12

US-10-108-260A-3849
Sequence 3849, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: No. US20040005560A1 full length cDNA
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3849
LENGTH: 693
TYPE: PRT
ORGANISM: Homo sapiens
US-10-108-260A-3849

Query Match 14.6%; Score 171.5; DB 12; Length 693;
Best Local Similarity 31.6%; Pred. No. 0.00015;
Matches 62; Conservative 17; Mismatches 78; Indels 39; Gaps 11;

QY 33 PLHPSLNIPIYGRNLPPPLYRRPVTVPVPGTYYDTGLPSYPMWLTSPGFYVYHNG 92
DB 253 VSPSATPP--SQAPPSLAAPPLQVPSPPASP-----PMSPSATPPQAPP----- 297
QY 93 FLAT-QLNVPLPPLPRGFPVPPSRFPFAAAPP-----AAPPIAAEPAAALTLTTPVA 145
DB 298 -PLAAPPLQVPPSPAPSPMS-----SATPPRVPLLAAPLQVPSPPASLPSMFLA 351
QY 146 AEPAAAPVAABPAABPAABPAABPAABPAABPAABPAABPAABPAABPAABPA 202
DB 352 KPPPQAPPLATPLQA-LPSPSPSPQAPFS--PSASLP--MSPLATPPQAPVTLA 406
QY 203 P---AAEPHSPSLE 215
DB 407 PLLQVPPSPSPSPTLQ 422

RESULT 13

US-10-084-843-142
Sequence 142, Application US/10084843
Publication No. US20030134243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Iodes, Michael J.
APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESSES:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9

Tue Jan 20 17:43:12 2004

us-09-923-236-2.rapb

Page 7

Db 348 A--PAPETKSDAEPAAAAAPPETKSDAAAAADPAFGTKADAA 387

Search completed: January 15, 2004, 15:19:47
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:11:59 ; Search time 21 Seconds
(without alignments)
441.242 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178

Sequence: 1 MKLLWACIVCAFARKRRP.....TAKPAAPRHPSPSLBOANQ 219

Scoring table: BLOSUM62

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/BACKFILE1.pep.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1178	100.0	219	US-09-527-345-2	Sequence 2, Appl
2	1179	15.2	207	US-07-836-642-2	Sequence 2, Appl
3	1179	15.2	207	US-08-169-563-2	Sequence 2, Appl
4	1179	15.2	207	US-08-403-379A-6	Sequence 6, Appl
5	1179	15.2	207	US-08-557-309B-50	Sequence 50, Appl
6	1179	15.2	207	US-08-929-414-6	Sequence 6, Appl
7	1179	15.2	207	US-08-834-306-50	Sequence 50, Appl
8	1179	15.2	207	US-08-993-674A-50	Sequence 50, Appl
9	1179	15.2	207	US-09-256-976-50	Sequence 50, Appl
10	1179	14.4	267	US-08-818-112-142	Sequence 142, App
11	1170	14.4	267	US-08-818-111-137	Sequence 137, App
12	1170	14.4	267	US-09-056-556-142	Sequence 142, App
13	1170	14.4	267	US-09-072-596-137	Sequence 137, App
14	164.5	14.0	316	US-09-252-991A-32957	Sequence 32957, A
15	163.5	13.9	805	US-09-103-429A-4	Sequence 4, Appl
16	154	13.1	805	US-09-103-429A-4	Sequence 4, Appl
17	153	13.0	786	US-09-103-429A-3	Sequence 3, Appl
18	150	12.7	399	US-09-252-991A-22853	Sequence 22853, A
19	148	12.6	2972	US-09-579-181-2	Sequence 2, Appl
20	148	12.6	3118	US-09-579-181-1	Sequence 1, Appl
21	145.5	12.4	1565	US-09-252-991A-20967	Sequence 20967, A
22	144	12.2	195	US-09-252-991A-27942	Sequence 27942, A
23	142.5	12.1	515	US-09-252-991A-25394	Sequence 25394, A
24	141	12.0	369	US-07-757-022B-14	Sequence 14, Appl
25	139.5	11.8	941	US-07-757-022B-84	Sequence 84, Appl
26	139.5	11.8	1022	US-07-757-022B-84	Sequence 84, Appl
27	139.5	11.8	1038	US-07-757-022B-74	Sequence 74, Appl

28	139.5	11.8	1049	US-07-757-022B-58	Sequence 58, Appl
29	139.5	11.8	1140	US-07-757-022B-104	Sequence 104, Appl
30	139.5	11.8	1270	US-07-757-022B-44	Sequence 44, Appl
31	139.5	11.8	1311	US-07-757-022B-42	Sequence 42, Appl
32	139.5	11.8	1313	US-07-757-022B-142	Sequence 142, Appl
33	139.5	11.8	1314	US-07-757-022B-50	Sequence 50, Appl
34	139.5	11.8	1320	US-07-757-022B-46	Sequence 46, Appl
35	139.5	11.8	1330	US-07-757-022B-60	Sequence 60, Appl
36	139.5	11.8	1354	US-07-757-022B-40	Sequence 40, Appl
37	139.5	11.8	1361	US-07-757-022B-52	Sequence 52, Appl
38	139.5	11.8	1363	US-07-757-022B-52	Sequence 52, Appl
39	139.5	11.8	1404	US-07-757-022B-2	Sequence 2, Appl
40	139.5	11.8	1404	US-07-757-022B-62	Sequence 62, Appl
41	138	11.7	180	US-07-757-022B-7	Sequence 7, Appl
42	138	11.7	180	US-07-757-022B-6	Sequence 6, Appl
43	138	11.7	304	US-09-252-991A-23116	Sequence 23116, A
44	136.5	11.6	504	US-09-219-849-3	Sequence 3, Appl
45	136.5	11.6	561	US-08-642-255-52	Sequence 52, Appl

ALIGNMENTS

```

RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; PRIOR FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match      100.0%; Score 1178; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 3,3e-93;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKLLWACIVCAFARKRRPFIQEDDNDGHLAPSLNIPYGRMLPPPLYRPNVTVP 60
DB      1 MKLLWACIVCAFARKRRPFIQEDDNDGHLAPSLNIPYGRMLPPPLYRPNVTVP 60

QY      61 SYGNTYTDGLPSYRWILTSQGFPPVYHIRGFLATQNLVPPLPGRGFPVPSRPSA 120
DB      61 SYGNTYTDGLPSYRWILTSQGFPPVYHIRGFLATQNLVPPLPGRGFPVPSRPSA 120

QY      121 AAPAPAPIAAPAPAPATATVPAAPAAAGVAAAPAAEAPVGAAPAAEAPAA 180
DB      121 AAPAPAPIAAPAPAPATATVPAAPAAAGVAAAPAAEAPVGAAPAAEAPAA 180

QY      181 EAPVGEPAAPAEPSAPAPATAPKAPADEPHSPSLBOANQ 219
DB      181 EAPVGEPAAPAEPSAPAPATAPKAPADEPHSPSLBOANQ 219

RESULT 2
US-07-836-642-2
; Sequence 2, Application US/07836642
; Patent No. 5304371
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
; TITLE OF INVENTION: Against T. cruzi Infection

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/836,642
FILING DATE: 14-FEB-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: REED
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-836-642-2

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;
QY 114 PSRFSAAAAPAPPIAEPAAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGVPEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 3
US-08-169-563-2
Sequence 2, Application US/08169563
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
TITLE OF INVENTION: Against T. cruzi Infection
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,563
FILING DATE: 17-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/836,642
FILING DATE: 14-FEB-1992
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: REED-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-563-2

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;
QY 114 PSRFSAAAAPAPPIAEPAAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGVPEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 4
US-08-403-379A-6
Sequence 6, Application US/08403379A
Patent No. 5756662
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6100 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,379A
FILING DATE: 14-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-379A-6

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;
QY 114 PSRFSAAAAPAPPIAEPAAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173

Db 5 PAEPKSAEPKPAEPK-SABEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
 Qy 174 VAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 213
 Db 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 5

US-08-557-3098-50
 ; Sequence 50, Application US/085573098
 ; Patent No. 5916572
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Houghton, Raymond L.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
 ; NUMBER OF SEQUENCES: 69
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/557,3098
 ; FILING DATE: 14-NOV-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.422
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 50:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 207 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-557-3098-50

Query Match 15.2%; Score 179; DB 2; Length 207;
 Best Local Similarity 41.0%; Pred. No. 4.1e-08;
 Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

Qy 114 PSRFFSAAAPPAAPPIAEPKAAAPLTATVAAPPAAGAVAAEPKAAEPVGAEPKAAEP 173
 Db 5 PAEPKSAEPKPAEPK-SABEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
 Qy 174 VAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 213
 Db 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 6

US-08-929-414-6
 ; Sequence 6, Application US/08929414
 ; Patent No. 5942403
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Haughton, Raymond
 ; APPLICANT: Skeiky, Yasir A.W.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
 ; TITLE OF INVENTION: OPT. CRUZI INFECTION
 ; NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/929,414
 ; FILING DATE: 15-SEP-1997
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.406C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 207 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-929-414-6

Query Match 15.2%; Score 179; DB 2; Length 207;
 Best Local Similarity 41.0%; Pred. No. 4.1e-08;
 Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

Qy 114 PSRFFSAAAPPAAPPIAEPKAAAPLTATVAAPPAAGAVAAEPKAAEPVGAEPKAAEP 173
 Db 5 PAEPKSAEPKPAEPK-SABEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
 Qy 174 VAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 213
 Db 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 7

US-08-834-306-50
 ; Sequence 50, Application US/08834306
 ; Patent No. 6054135
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Houghton, Raymond L.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
 ; NUMBER OF SEQUENCES: 65
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/834,306
 ; FILING DATE: 15-APR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-50

Query Match 15.2%; Score 179; DB 3; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSHFSAAPAAPAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 8
US-08-993-674A-50
Sequence 50, Application US/08993674A
Patent No. 6228372

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-50

Query Match 15.2%; Score 179; DB 3; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSHFSAAPAAPAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 9
US-09-256-976-50
Sequence 50, Application US/09256976
Patent No. 6419933

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
FILE REFERENCE: 210121.422C3
CURRENT APPLICATION NUMBER: US/09/256,976
CURRENT FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 50
LENGTH: 207
TYPE: PR1
ORGANISM: Trypanosoma cruzi
US-09-256-976-50

Query Match 15.2%; Score 179; DB 4; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSHFSAAPAAPAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 10
US-08-818-112-142
Sequence 142, Application US/08818112
Patent No. 6290969

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

RESULT 11
 US-08-618-111-137
 ; Sequence 137, Application US/06818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skelky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonia
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzlk, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 NUMBER OF SEQUENCES: 148
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 City: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/618,111
 FILING DATE: 13-MAR-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:

RESULT 12
 US-09-056-556-142
 Sequence 142, Application US/09056556
 Patent No. 6350436
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.M.
 APPLICANT: Dillon, Davin C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 NUMBER OF SEQUENCES: 241
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-APR-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 142:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 267 amino acids
 TYPE: amino acid


```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-217-327-4

```

Query Match 13.9%; Score 163.5; DB 1; Length 214;

Best Local Similarity 30.3%; Pred. No. 8.9e-07;

Matches 67; Conservative 20; Mismatches 81; Indels 53; Gaps 14;

```

QY 2 KLLIMACIVCAFAKRRPPFIGEDNDGHPHAPSINIPYGINRLPPPLYRPPVNTVPS 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 KNLFLSALLCIAVA-----GVLGQ-----APSNPRTSTP----ATPTP----PASTPPP 46
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 YPGNTYTDGLPSYFWILTSPPGFYVYHNGFPLATQLNVPL---PPRGFPFVPPSRF 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 TTAAPPPTIATP--PVTSTP-----PTSSP---PVTASPPVSTP--PPS-- 86
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 FSAALAAPAPPIAEPAAALPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAE 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 ---SPPATPPPPASPPATPPPPASPPATPPPPASPPATPPPPATPPPPATPPPA-- 141
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 178 PAEAPVGEPAABEPSPAE-PATAKPAEPHPSPSLAQ 217
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 142 PLASPPATV-PAT---SPVQTPLTSPAPPTAAPTIGAA 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: January 15, 2004, 15:15:17
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:29:52 ; Search time 33 Seconds

(without alignments)
1357.012 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MKLLWACIVCAVAFARRRRF.....TAKRAPEHPSPLEQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 762491 seqs, 204481190 residues

Word size: 0

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	9 US-09-922-480-2	Sequence 2, Appli
2	219	100.0	219	9 US-09-923-236-2	Sequence 2, Appli
3	219	100.0	219	10 US-09-922-469-2	Sequence 2, Appli
4	150	68.5	221	12 US-10-231-417-192	Sequence 192, App
5	140	63.9	219	11 US-09-746-783-106	Sequence 106, App
6	9	4.1	798	15 US-10-156-761-13162	Sequence 13162, A
7	9	4.1	2969	10 US-09-738-626-4434	Sequence 4434, Ap
8	9	3.7	132	12 US-10-289-762-96	Sequence 96, Appl
9	9	3.7	206	15 US-10-219-220-250	Sequence 250, App
10	9	3.7	249	15 US-10-180-375-82	Sequence 82, Appl
11	9	3.7	342	12 US-10-369-493-12383	Sequence 12383, A
12	9	3.7	367	12 US-10-369-493-7317	Sequence 7317, Ap
13	9	3.7	383	10 US-09-788-345-10	Sequence 10, Appl
14	9	3.7	383	12 US-10-337-312-10	Sequence 12, Appl
15	9	3.7	412	10 US-09-788-345-12	Sequence 12, Appl

16	8	3.7	412	12 US-10-337-312-12	Sequence 12, Appl
17	8	3.7	428	10 US-09-906-514-4	Sequence 4, Appli
18	8	3.7	428	12 US-10-177-478-14	Sequence 14, Appli
19	8	3.7	454	15 US-10-156-761-13939	Sequence 13939, A
20	8	3.7	511	9 US-09-864-761-34590	Sequence 34590, A
21	8	3.7	545	12 US-10-929-386-32280	Sequence 32280, A
22	8	3.7	695	12 US-10-369-493-15364	Sequence 15364, A
23	8	3.7	850	12 US-10-369-493-15364	Sequence 10111, A
24	8	3.7	4307	12 US-10-369-493-5698	Sequence 5698, Ap
25	8	3.7	4307	12 US-10-369-493-5698	Sequence 5698, Ap
26	8	3.7	4307	12 US-10-369-493-5698	Sequence 5700, Ap
27	8	3.7	8026	12 US-10-132-134-12	Sequence 12, Appl
28	8	3.2	16	12 US-10-397-551-43	Sequence 43, Appl
29	8	3.2	21	14 US-10-124-800-15	Sequence 15, Appl
30	7	3.2	27	11 US-09-974-879-306	Sequence 306, App
31	7	3.2	27	11 US-09-305-736-306	Sequence 306, App
32	7	3.2	27	11 US-09-818-683-306	Sequence 306, App
33	7	3.2	48	9 US-09-739-907-177	Sequence 177, App
34	7	3.2	48	12 US-09-938-671-177	Sequence 177, App
35	7	3.2	60	9 US-09-864-761-33554	Sequence 33554, A
36	7	3.2	60	9 US-09-864-761-33624	Sequence 33624, A
37	7	3.2	60	9 US-09-864-761-34127	Sequence 34127, A
38	7	3.2	61	9 US-09-864-761-39187	Sequence 39187, A
39	7	3.2	65	9 US-09-864-761-36194	Sequence 36194, A
40	7	3.2	78	11 US-09-820-843A-89	Sequence 89, Appl
41	7	3.2	88	9 US-09-764-853-503	Sequence 503, App
42	7	3.2	91	9 US-09-867-550-738	Sequence 738, App
43	7	3.2	95	9 US-09-739-907-178	Sequence 178, App
44	7	3.2	95	12 US-09-938-671-178	Sequence 178, App
45	7	3.2	105	12 US-10-104-047-2874	Sequence 2874, Ap

ALIGNMENTS

RESULT 1
US-09-922-480-2
Sequence 2, Application US/09922480
Patent No. US20020081701A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OR INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/922,480
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-480-2

Query Match	100.0%	Score 219	DB 9	Length 219
Best Local Similarity	100.0%	Pred. No. 5.6e-185		
Matches 219	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MKLLWACIVCAVAFARRRRFPGIGEDNDGHPHLSLNTPYGIRNLPPLYYRPNVTVP	60	
DB	1	MKLLWACIVCAVAFARRRRFPGIGEDNDGHPHLSLNTPYGIRNLPPLYYRPNVTVP	60	
QY	61	SYRNTYTDGLSGYPVILTSPPGPPYHHRGPPPLAQVLNVPPLPPGPFVPSRRFSA	120	
DB	61	SYRNTYTDGLSGYPVILTSPPGPPYHHRGPPPLAQVLNVPPLPPGPFVPSRRFSA	120	
QY	121	AAAPAPPIAEPAAAPLFTATPAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA	180	
DB	121	AAAPAPPIAEPAAAPLFTATPAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAA	180	
QY	181	EAPVGAEPAAEPSPAPATAKRAPEHPSPLEQANQ	219	

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

RESULT 2
US-09-923-236-2
; Sequence 2, Application US/099223236

; Patent No. US20020090677A1
; GENERAL INFORMATION:

; APPLICANT: Sheppard, Paul O.

; APPLICANT: Adler, David A.

; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE

; FILE REFERENCE: 97-71

; CURRENT APPLICATION NUMBER: US/09/923,236

; PRIOR FILING DATE: 2001-08-03

; PRIOR APPLICATION NUMBER: US 60/124,820

; NUMBER OF SEQ. ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 219

; TYPE: PRP

; ORGANISM: Homo sapiens

US-09-923-236-2

Query Match 100.0%; Score 219; DB 9; Length 219;

Best Local Similarity 100.0%; Pred. No. 5.6e-185;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACTVCVAFARRRPFI GEDDNDGHLPSLNI PYGIRNLPPPLYRRPVNTVP 60

Db 1 MKLLMACTVCVAFARRRPFI GEDDNDGHLPSLNI PYGIRNLPPPLYRRPVNTVP 60

QY 61 SYGNTYTDGTPSYWILTSRGPYYVYHIRGFPLATQUNVPLPRGPFVPSRFFSA 120

Db 61 SYGNTYTDGTPSYWILTSRGPYYVYHIRGFPLATQUNVPLPRGPFVPSRFFSA 120

QY 121 AAAPAPPIAEPAAAAPLTATPVAAEPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180

Db 121 AAAPAPPIAEPAAAAPLTATPVAAEPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 61 SYGNTYTDGTPSYWILTSRGPYYVYHIRGFPLATQUNVPLPRGPFVPSRFFSA 120

Db 61 SYGNTYTDGTPSYWILTSRGPYYVYHIRGFPLATQUNVPLPRGPFVPSRFFSA 120

QY 121 AAAPAPPIAEPAAAAPLTATPVAAEPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180

Db 121 AAAPAPPIAEPAAAAPLTATPVAAEPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

QY 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATKAPAEPPHPSLEQANQ 219

```

; Fechtel, Kim
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: Genetics Institute, Inc.
; STREET: 87 Cambridge Park Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/746,783
; FILING DATE: 21-Dec-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mlaalincic, Debra J.
; REGISTRATION NUMBER: 46,931
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-09-746-783-106

Query Match      63.9%; Score 140; DB 11; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.1e-115;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKLLIMACTIVCAARKRRPFPIGDDNDGHPHPSLNTYIGIRNLPPPLYTPPVNTVP 60
DB      1 MKLLIMACTIVCAARKRRPFPIGDDNDGHPHPSLNTYIGIRNLPPPLYTPPVNTVP 60
QY      61 SYPGNTYDTGLPSYPMILTSRGPYVYHIRGPPATQNLNPPPLPPRGFPVPSRPFSA 120
DB      61 SYPGNTYDTGLPSYPMILTSRGPYVYHIRGPPATQNLNPPPLPPRGFPVPSRPFSA 120
QY      121 AAAPAPPIAAEPAAAPLT 140
DB      121 AAAPAPPIAAEPAAAPLT 140

RESULT 6
US-10-156-761-13162
; Sequence 13162, Application US/10156761
; Publication No. US20030113018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

```

```

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13162
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13162

Query Match      4.1%; Score 9; DB 15; Length 798;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      198 PATAKPAAP 206
DB      29 PATAKPAAP 37

RESULT 7
US-09-738-626-4434
; Sequence 4434, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIJO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4434
; LENGTH: 2969
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4434

Query Match      4.1%; Score 9; DB 10; Length 2969;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      168 PAAEAPVVA 176
DB      1663 PAAEAPVVA 1671

RESULT 8
US-10-289-762-96
; Sequence 96, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; PRIOR FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 96
; LENGTH: 132

```

TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-10-289-762-96

Query Match 3.7%; Score 8; DB 12; Length 132;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAEP 178
DB 53 EAPVAEP 60

RESULT 9
US-10-219-220-250
Sequence 250, Application US/10219220
Publication No. US20030082724A1
GENERAL INFORMATION:
APPLICANT: Flinn, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: Compositions affecting programmed cell
death and their use in the modification of plant development
FILE REFERENCE: 11000.102261
CURRENT APPLICATION NUMBER: US/10/219,220
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
PRIOR FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 250
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 250
LENGTH: 206
TYPE: PRT
ORGANISM: Eucalyptus grandis
US-10-219-220-250

Query Match 3.7%; Score 8; DB 15; Length 206;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 112 AAAAPAP 119

RESULT 10
US-10-180-375-82
Sequence 82, Application US/10180375
Publication No. US2003012638A1
GENERAL INFORMATION:
APPLICANT: Allen, William B.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Famodu, Omolayo O.
APPLICANT: Harvell, Leslie T.
APPLICANT: Helentjaris, Timothy
APPLICANT: Li, Changliang
APPLICANT: Lowe, Keith
APPLICANT: Oliveira, Igor Cunha
APPLICANT: Shen, Bo
APPLICANT: Tarczyński, Mitchell C.
TITLE OF INVENTION: Alteration of oil traits in plants
FILE REFERENCE: B01458 US NAI
CURRENT APPLICATION NUMBER: US/10/180,375
PRIOR FILING DATE: 2002-06-26
NUMBER OF SEQ ID NOS: 222
SOFTWARE: Microsoft Office 97
SEQ ID NO 82
LENGTH: 249
TYPE: PRT
ORGANISM: Trifolium aestivum
US-10-180-375-82

Query Match 3.7%; Score 8; DB 15; Length 249;
Best Local Similarity 100.0%; Pred. No. 42;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 VEPAAEP 193
DB 92 VEPAAEP 99

RESULT 11
US-10-369-493-12383
Sequence 12383, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianning
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12383
LENGTH: 342
TYPE: PRT
ORGANISM: Aspergillus nidulans
FEATURE:
NAME/KEY: unsure
LOCATION: (1) .. (342)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12383

Query Match 3.7%; Score 8; DB 12; Length 342;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 126
DB 297 SAAAPAA 304

RESULT 12
US-10-369-493-7317
Sequence 7317, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianning
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 7317
LENGTH: 367
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-369-493-7317

Query Match 3.7%; Score 8; DB 12; Length 367;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156

Db 71 AAGAPVAA 78

```

RESULT 13
US-09-788-345-10
; Sequence 10, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-09-788-345-10

```

Query Match 3.7%; Score 8; DB 10; Length 383;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 354 AAAAPAP 361

```

RESULT 14
US-10-337-312-10
; Sequence 10, Application US/10337312
; Publication No. US20030138451A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/10/337,312
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US/09/788,345
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-10-337-312-10

```

Query Match 3.7%; Score 8; DB 12; Length 383;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127

Db 354 AAAAPAP 361

```

RESULT 15
US-09-788-345-12
; Sequence 12, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant sequence from vectors pQ31 and pMal
US-09-788-345-12

```

Query Match 3.7%; Score 8; DB 10; Length 412;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 383 AAAAPAP 390

Search completed: January 15, 2004, 15:35:27
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:28:37 ; Search time 21 Seconds
(without alignments)
441.242 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MKLLIMACIVCAFAFKRRF.....TAKPAAPRHPSPSLEQNMQ 219

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	US-09-527-345-2	Sequence 2, Appli
2	9	4.1	322	US-08-428-414A-2	Sequence 2, Appli
3	8	3.7	132	US-09-198-452A-96	Sequence 96, Appli
4	8	3.7	168	US-09-252-991A-1819	Sequence 1819, A
5	8	3.7	190	US-09-252-991A-17963	Sequence 17963, A
6	8	3.7	307	US-09-252-991A-21588	Sequence 21588, A
7	8	3.7	383	US-09-471-396-3	Sequence 3, Appli
8	8	3.7	383	US-09-788-345-10	Sequence 10, Appli
9	8	3.7	410	US-09-252-991A-30606	Sequence 30606, A
10	8	3.7	412	US-09-471-396-1	Sequence 1, Appli
11	8	3.7	412	US-09-788-345-12	Sequence 12, Appli
12	8	3.7	549	US-09-252-991A-20196	Sequence 20196, A
13	8	3.7	575	US-08-653-648A-5	Sequence 5, Appli
14	8	3.7	778	US-09-252-991A-27591	Sequence 27591, A
15	8	3.7	4545	US-08-804-227C-14	Sequence 14, Appli
16	8	3.7	4550	US-08-804-227C-8	Sequence 8, Appli
17	8	3.7	4550	US-08-804-198-2	Sequence 2, Appli
18	19	7.3	31	US-09-248-588-27	Sequence 27, Appli
19	19	7.3	76	5273901-11	Patent No. 5273901
20	20	7.3	76	5482709-10	Patent No. 5482709
21	21	7.3	76	5496550-6	Patent No. 5496550
22	22	7.3	79	US-09-252-991A-25964	Sequence 25964, A
23	23	7.3	118	US-08-301-162-10	Sequence 10, Appli
24	24	7.3	118	US-09-461-240-10	Sequence 10, Appli
25	25	7.3	118	US-09-968-927-10	Sequence 10, Appli
26	26	7.3	122	US-09-328-352-7934	Sequence 7934, A
27	27	7.3	134	US-09-252-991A-18886	Sequence 18886, A

28	7	3.2	136	4	US-09-732-210-814	Sequence 814, App
29	7	3.2	136	4	US-09-732-210-815	Sequence 815, App
30	7	3.2	141	4	US-09-252-991A-26122	Sequence 26122, A
31	7	3.2	156	1	US-08-074-121-5	Sequence 5, Appli
32	7	3.2	156	3	PCT-US94-06447-5	Sequence 13, Appli
33	7	3.2	159	3	US-08-481-435-13	Sequence 17069, A
34	7	3.2	170	4	US-09-252-991A-27069	Sequence 17373, A
35	7	3.2	173	4	US-09-252-991A-17373	Sequence 25321, A
36	7	3.2	173	4	US-09-252-991A-25321	Patent No. 5273901
37	7	3.2	180	6	5482709-6	Patent No. 5482709
38	7	3.2	188	4	US-09-252-991A-20399	Sequence 20399, A
39	7	3.2	195	4	US-09-252-991A-20967	Sequence 20967, A
40	7	3.2	197	4	US-09-252-991A-24218	Sequence 24218, A
41	7	3.2	202	4	US-09-252-991A-28326	Sequence 28326, A
42	7	3.2	216	4	US-09-252-991A-18761	Sequence 18761, A
43	7	3.2	239	4	US-09-252-991A-21250	Sequence 21250, A
44	7	3.2	249	4	US-09-252-991A-18733	Sequence 18733, A
45	7	3.2	249	4	US-09-252-991A-18733	Sequence 18733, A

ALIGNMENTS

```
RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match      100.0%; Score 219; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.1e-203; Indels 0; Gaps 0;
Matches 219; Conservative 0; Mismatches 0;

QY      1 MKLLIMACIVCAFAFKRRFPFEGEDNDGHPHPSINTPYGIRNLPPPLYRPPVNTVP 60
      |||||
DB      1 MKLLIMACIVCAFAFKRRFPFEGEDNDGHPHPSINTPYGIRNLPPPLYRPPVNTVP 60
      |||||

QY      61 SYPNNTYTDGTPSYPMILTSFGFPYVYHKGPLATQNLVPLPRGFPVPPSRRFSA 120
      |||||
DB      61 SYPNNTYTDGTPSYPMILTSFGFPYVYHKGPLATQNLVPLPRGFPVPPSRRFSA 120
      |||||

QY      121 AAAAAPPAAEPAAAPLRTATPVAABPAAGAPVAAABPAABPAABPAABPAABPAAB 180
      |||||
DB      121 AAAAAPPAAEPAAAPLRTATPVAABPAAGAPVAAABPAABPAABPAABPAABPAAB 180
      |||||

QY      181 EAPVGEPPAAEPSPAPATAPKPAAPRHPSPSLEQNMQ 219
      |||||
DB      181 EAPVGEPPAAEPSPAPATAPKPAAPRHPSPSLEQNMQ 219
      |||||

RESULT 2
US-08-428-414A-2
; Sequence 2, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; LEISHMANIASIS
```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecsek, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-428-414A-2

Query Match 4.1%; Score 9; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAP 138
Db 289 AAEPAAAP 297

RESULT 3
US-09-198-452A-96
Sequence 96, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 96
LENGTH: 132
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-96

Query Match 3.7%; Score 8; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAAP 178
Db 53 EAPVAAAP 60

RESULT 4
US-09-252-991A-18199
Sequence 18199, Application US/09252991A

Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18199
LENGTH: 168
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18199

Query Match 3.7%; Score 8; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 VAAEPAAAG 151.
Db 108 VAAEPAAAG 115

RESULT 5
US-09-252-991A-17963
Sequence 17963, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17963
LENGTH: 190
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17963

Query Match 3.7%; Score 8; DB 4; Length 190;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAAAP 128
Db 113 AAAPAAAP 120

RESULT 6
US-09-252-991A-21588
Sequence 21588, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 21588
 ; LENGTH: 307
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-21588

Query Match
 Best Local Similarity 100.0%; Score 8; DB 4; Length 307;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAAPAP 128
 Db 120 AAAAPAP 127

RESULT 7
 US-09-471-396-3
 ; Sequence 3, Application US/09471396
 ; Patent No. 6458359
 ; GENERAL INFORMATION:
 ; APPLICANT: BEDATE, Carlos Alonso
 ; APPLICANT: REQUEENA ROLANTA, Jose Maria
 ; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
 ; TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
 ; TITLE OF INVENTION: L. INFANTUM AND PROTEIN ENCODED BY SAIDI GENE, AND ...
 ; FILE REFERENCE: bedate2a.seq
 ; CURRENT APPLICATION NUMBER: US/09/471,396
 ; CURRENT FILING DATE: 1999-12-23
 ; PRIOR APPLICATION NUMBER: 60/113,825
 ; PRIOR FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 383
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Protein Q
 US-09-471-396-3

Query Match
 Best Local Similarity 100.0%; Score 8; DB 4; Length 383;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 354 AAAAPAP 361

RESULT 8
 US-09-788-345-10
 ; Sequence 10, Application US/09788345
 ; Patent No. 6525186
 ; GENERAL INFORMATION:
 ; APPLICANT: ALONSO BEDATE, Carlos
 ; APPLICANT: REQUEENA ROLANTA, Jose M.
 ; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
 ; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
 ; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
 ; FILE REFERENCE: BEDATE-1A
 ; CURRENT APPLICATION NUMBER: US/09/788,345
 ; CURRENT FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: 09/219,306
 ; PRIOR FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: Patentln version 3.1
 ; SEQ ID NO 10
 ; LENGTH: 383
 ; TYPE: PRT

; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chimeric
 US-09-788-345-10

Query Match
 Best Local Similarity 100.0%; Score 8; DB 4; Length 383;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 354 AAAAPAP 361

RESULT 9
 US-09-252-991A-30606
 ; Sequence 30606, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 30606
 ; LENGTH: 410
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-30606

Query Match
 Best Local Similarity 100.0%; Score 8; DB 4; Length 410;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVVA 156
 Db 37 AAGAPVVA 44

RESULT 10
 US-09-471-396-1
 ; Sequence 1, Application US/09471396
 ; Patent No. 6458359
 ; GENERAL INFORMATION:
 ; APPLICANT: BEDATE, Carlos Alonso
 ; APPLICANT: REQUEENA ROLANTA, Jose Maria
 ; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
 ; TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
 ; TITLE OF INVENTION: L. INFANTUM AND PROTEIN ENCODED BY SAIDI GENE, AND ...
 ; FILE REFERENCE: bedate2a.seq
 ; CURRENT APPLICATION NUMBER: US/09/471,396
 ; CURRENT FILING DATE: 1999-12-23
 ; PRIOR APPLICATION NUMBER: 60/113,825
 ; PRIOR FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 412
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Protein Q
 US-09-471-396-1

Query Match
 Best Local Similarity 100.0%; Score 8; DB 4; Length 412;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
|||||

DB 383 AAAAPAP 390

RESULT 11

US-09-788-345-12
; Sequence 12, Application US/09788345

; Patent No. 6525186

; GENERAL INFORMATION:

; APPLICANT: ALONSO BEDATE, Carlos

; APPLICANT: ROSUENA ROLANLA, Jose M.

; APPLICANT: SOTO ALVAREZ, Manuel

; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN

; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG

; FILE REFERENCE: BEDATE=1A

; CURRENT APPLICATION NUMBER: US/09/788,345

; PRIOR FILING DATE: 2001-02-21

; PRIOR APPLICATION NUMBER: 09/219,306

; PRIOR FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 12

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURES:

; OTHER INFORMATION: recombinant sequence from vectors pQ31 and pMal

US-09-788-345-12

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 412;
Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
|||||

DB 383 AAAAPAP 390

RESULT 12

US-09-252-991A-20196
; Sequence 20196, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20196

; LENGTH: 549

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20196

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 549;
Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 AAEPAAGA 152
|||||

DB 326 AAEPAAGA 333

RESULT 13
US-08-653-648A-5
; Sequence 5, Application US/08653648A

; Patent No. 6379945

; GENERAL INFORMATION:

; APPLICANT: Jepson, Ian

; APPLICANT: Greenland, Andrew

; APPLICANT: Martinez, Alberto

; TITLE OF INVENTION: A Gene Switch

; FILE REFERENCE: PPD50047/US

; CURRENT APPLICATION NUMBER: US/08/653,648A

; PRIOR FILING DATE: 1996-05-24

; PRIOR APPLICATION NUMBER: GB 9510759.5

; PRIOR FILING DATE: 1995-05-26

; PRIOR APPLICATION NUMBER: GB 9605656.9

; PRIOR FILING DATE: 1996-03-18

; PRIOR APPLICATION NUMBER: GB 9513882.2

; PRIOR FILING DATE: 1995-07-07

; PRIOR APPLICATION NUMBER: GB 9517316.7

; PRIOR FILING DATE: 1995-08-24

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 5

; LENGTH: 575

; TYPE: PRT

; ORGANISM: Helicobacter virescens

US-08-653-648A-5

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 575;
Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 TATPVAAE 147
|||||

DB 552 TATPVAAE 559

RESULT 14

US-09-252-991A-27591
; Sequence 27591, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27591

; LENGTH: 778

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27591

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 778;
Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 PAAGAPVA 155
|||||

DB 655 PAAGAPVA 662

RESULT 15

US-08-804-227C-14
; Sequence 14, Application US/08804227C

; Patent No. 5876991

; GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sulton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-14

Query Match 3.7%; Score 8; DB 2; Length 4545;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 131 AEPAAAP 138
DB 2620 AEPAAAP 2627

Search completed: January 15, 2004, 15:34:05
Job time : 21 secs